YALE ECONOMIC ESSAYS

VOLUME 3 NUMBER 1

Determinants of Bond Yield Differentials—1954–1959 *Peter E. Sloane*

Monetary Analysis in an Underdeveloped Economy: A Case Study of Three East African Territories *Donald C. Mead*

A Multicountry Model of Forward Exchange: Theory, Policy, and Empirical Evidence, 1955–1961 *Herbert G. Grubel*

Wage Changes in 24 Manufacturing Industries, 1948–1959: A Comparative Analysis *Donald R. Snodgrass*

Exclusive Dealing in the Petroleum Industry: The Refiner-Lessee Dealer Relationship *Richard A. Miller*

SPRING 1963

YALE ECONOMIC ESSAYS

This journal is intended to serve two principal functions. All doctoral dissertations in economics at Yale University will be published in its pages either in full or in extended summary. It is also a medium for economic comment and analysis by faculty, students, and graduates in economics of Yale University.

Editorial Board: Neil W. Chamberlain
Raymond P. Powell Arthur M. Okun

Volume 3 Spring 1963 Number 1

Yale Economic Essays is printed at the Carl Purington Rollins Printing-Office of the Yale University Press. A one-year subscription (two issues) is $4.00. Single copies are $2.50. The rate for students for one year is $2.00. Checks should be made payable to Yale University. Please address any correspondence to the Editorial Board, 701 Sterling-Sheffield-Strathcona Hall, New Haven, Connecticut.

Copyright 1963 by the Yale Economic Essays.

YALE ECONOMIC ESSAYS

TABLE OF CONTENTS

Determinants of Bond Yield Differentials—1954–1959 *Peter E. Sloane* 3

Monetary Analysis in an Underdeveloped Economy: A Case Study of Three East African Territories *Donald C. Mead* 57

A Multicountry Model of Forward Exchange: Theory, Policy, and Empirical Evidence, 1955–1961 *Herbert G. Grubel* 105

Wage Changes in 24 Manufacturing Industries, 1948–59: A Comparative Analysis *Donald R. Snodgrass* 171

Exclusive Dealing in the Petroleum Industry: The Refiner–Lessee Dealer Relationship *Richard A. Miller* 223

Summary: The purpose of this essay is to estimate empirically the risk premiums existing between substitutable assets of equal maturity at various points along their yield curves.

Traditionally both theoretical and empirical analyses of the structure of interest rates have been concerned with explaining the difference between yields on short and long term assets. This study utilizes bond market yield data for the years 1954 through 1959 to construct yield versus maturity curves for specific bond quality categories. With the U.S. Treasury bond curve serving as the basic yield curve, it is then possible to calculate the absolute yield differentials at various maturities. This risk premium serves as the dependent variable for the regression analysis undertaken. The results support those advocates of an expectational, as opposed to a loanable funds, approach to any adequate explanation of the over-all interest rate structure. It is further concluded that the shape of the long term yield curve can be closely approximated by a smooth, continuous curve, i.e., the yield curve is not comprised of a series of discontinuous yields based upon the specific conditions existing at each maturity.

This essay was submitted to Yale University in 1962 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics at Smith College.

Dissertation Committee: Professors Arthur Okun (Chairman), Harold Watts, and Gustav Ranis.

*

The topic was originally suggested and initial encouragement provided by Professor James Tobin. The carrying out and completion of the study relied heavily upon the good offices of the members of my committee, Professors Arthur M. Okun and Harold W. Watts. I should like to express my sincerest appreciation for continued interest in spite of the press of other duties and for their invaluable comments and suggestions.

I should also like to thank my colleagues Donald Hester, Donald Snodgrass, and T. N. Srinivasan for their advice and opinions on specific points.

The Cowles Foundation has financed all the computational work undertaken. A final word of thanks should go to the staff of the Yale Computing Center; they maintain an extreme degree of patience in face of the never ending stream of novices which flows through the portals.

P.E.S.

DETERMINANTS OF BOND YIELD DIFFERENTIALS
1954-1959

Peter E. Sloane

1. Introduction

The theory of the structure of interest rates has traditionally been confined to an analysis of the difference between yields on short term and long term assets. This empirical study deals with differential yields at specific, long term maturities, but, in spite of this fact, the same general theoretical problems are encountered. Actually, it is not until uncertainty is introduced that the relevance of traditional theory is evident. With riskless securities and perfect certainty there would be no room for a study of risk premiums among assets of different quality, i.e., default risk. Once these conditions are relaxed, the subjective estimation of uncertain future conditions becomes equally paramount in determining whether the risk of holding longer term securities should be undertaken and in the choice among assets of equal maturity but which are subject to varying degrees of risk with respect to interest and principal. Both decisions involve the confidence with which expectations are held. Choosing a longer term asset involves the possibility that anticipations will not be realized and that on this account a loss will be taken; choosing a lower quality asset involves the same type of probability. In either case the investor will demand a premium to make an investment the outcome of which is uncertain. A brief review of the existing theory, therefore, should serve as a useful background for consideration of the statistical analysis to follow.

In both the nonmonetary theory of Irving Fisher and the monetary theory of J. R. Hicks a first approximation of the long term rate is presented as the arithmetic average of the current and any anticipated, intervening short term rates. Abstracting from default risk, this assumption allows concentration upon the rate of interest since, once two or more short rates are known with certainty or with given expectations, the long rate can readily be calculated.[1] (As Hicks points out, the known and unknown rate can just as easily be reversed.[2]) In his "second approximation" Fisher allows variation in income streams due to size, composition and time shape, but not until the "third approximation" does he allow for uncertainty. His partiality to neatly posited theory under controlled conditions is evident in his conclusion: "The second approximation gives a clear cut theory applicable to the clear cut hypotheses on which it is based. The third approximation cannot avoid some degree of vagueness."[3] Hicks, building his basic theory on the relative

---

1. Irving Fisher, The Theory of Interest (reprint; New York: Kelley & Millman, Inc., 1954), p. 70. J. R. Hicks, Value and Capital (London: Oxford University Press, 1939), pp. 145-46.

2. Hicks, op. cit., p. 260.

3. Fisher, op. cit., p. 227.

"moneyness" of various assets, attributes that part of interest paid beyond "pure" interest to default risk and uncertainty concerning the future course of interest rates.[1] The elasticity of expectations is introduced as a possible key element in the determination of the course of interest rates. However, most of the ensuing discussion is in terms of quite inelastic expectations, since Hicks concludes that this condition is usual.[2]

F. A. Lutz, in a synthesis of the theory and an extension of the implications of expectations with respect to the interest rate structure,[3] builds a model, initially, under conditions of (1) certainty, (2) zero transactions costs and (3) perfect shiftability. Dropping the last two assumptions, Lutz draws certain implications concerning the relationship of rates for borrowers and lenders before turning to the consideration of risk. Abstracting from default risk, he considers the risk associated with changes in interest rates. He concludes that there is no precision in calling the long rate an average of expected short rates, but rather that the relationship is a highly complex one bearing upon the whole pattern of expectations held by many different people operating in the market. The implication is that the greater stability of the long over the short rate cannot be explained simply by the fact that a change in the single, current short rate will have less and less effect on the average as the lengthening of the term under consideration involves the averaging of more and more short rates. He is, nevertheless, able to justify this characteristic of the long rate by an argument, very close to Hicks' inelastic expectations, which states that the market has no very definite expectations concerning future short rates which are well into the future. A second conclusion is that the over-all rate structure may exhibit either a rising, falling or horizontal trend depending upon whether short rates are expected to rise, fall or remain the same.

In a more recent article, J. M. Culbertson[4] attacks the expectational theories of Hicks and Lutz, resting his case upon the demand for and the supply of funds. "The behaviour of rate structure does not seem explainable in terms

---

1. Hicks, op. cit., p. 163.

2. Hicks, op. cit., p. 282. The course of the statistical analysis will show how important this concept of elasticity is in explaining periods which are not normal and therefore may be given to a high degree of elasticity with respect to interest rate expectations.

3. F. A. Lutz, "The Structure of Interest Rates," Quarterly Journal of Economics (November 1940), pp. 36-63; reprinted in American Economic Association, Readings in the Theory of Income Distribution (Philadelphia: The Blakiston Company, 1946), Chap. 26, pp. 499-529.

4. J. M. Culbertson, "The Term Structure of Interest Rates," Quarterly Journal of Economics (November 1957), pp. 485-517. See also L. S. Wehrle, "Culbertson on Interest Structure: Comment," and J. M. Culbertson, "Reply," Quarterly Journal of Economics (November 1958), pp. 601-13.

of long run expectations, though near-term expectations can temporarily govern the behaviour of rates."[1] He identifies demand factors as new issues, shifts in purposes of borrowing, Treasury debt management and supply factors as shifts in liquidity, relative lending costs, lender expectations, patterns of speculative activity, changes in monetary policy, etc. Again, the discussion centers around whether short rates exceed long rates and why, although interest rate interrelationships are also mentioned. If we have a world of perfect certainty or if expectations are inelastic, Culbertson's criticisms and conclusions are perhaps justified, but if the relationships he discusses do, in themselves, affect expectations throughout the interest rate structure, it is hardly wrong to assign high importance to the so-called expectational theories.

J. W. Conard has done just this in his recent neo-classical synthesis.[2] Conard summarizes the implications of neo-classical theory of the interest rate structure in a world of certainty and perfect capital markets as: "for any defined period of time, however long or short, the expected net effective yield (including capital gains and losses) will be identical on all riskless securities regardless of their term."[3] With respect to expectations, neo-classical theory does not imply simple averaging of short rates but rather that investors seek to maximize returns on the basis of their expectations about future interest rates and changes in capital values. Only by adding considerations of risk and institutional factors can this theory be made realistic.

> The most fundamental fact which forces modification of the neoclassical theory is that of uncertainty. Some of the important consequences are these: (1) Expectations about conditions in the market for money and credit for the intermediate and long-term future are typically very vague or almost entirely unformed. Thus, long rates may be largely influenced by changes in short-term expectations and by considerations not described in neoclassical theory. (2) One of these other considerations is the fact that long-term securities offer more prospect of capital gain or loss than do shorts. The essential reason is that a simultaneous change in all expected short rates will cause a much larger change in long prices than in short prices, and a general change in expectations tends often to color the far as well as the immediate future, thus causing a similar if not an equal alteration in a substantial sequence of expected short rates. (3) A consequence of the more speculative character of longs is the tendency of institutions

---

1. Culbertson, "The Term Structure of Interest Rates," op. cit., p. 515.

2. J. W. Conard, Introduction to the Theory of Interest (Berkeley: University of California Press, 1959). It is interesting to note, as will be amplified later, that neither empirical results of this study nor of Conard's tend to support the demand for and supply of loanable funds as being the dominant factors in determining interest rate structure. See Conard, ibid., p. 338.

3. Ibid., p. 298

> requiring a high degree of liquidity (e.g., banks and nonfinancial corporations) to hold very few longs. (4) Another consequence is that longs often yield more than shorts, partly as compensation for risk-taking. (5) A result of 3 and 4 is that the markets for securities of different term become partly segmented. An important consequence for the theory of interest is that there is no longer a single rate of interest even defined over short periods of time as there was in the unmodified theory. Thus securities of different term are no longer perfect substitutes for one another, and a rate determined by the theory of interest is no longer unambiguous.[1]

The actual modification of the theory to allow for uncertainty and institutional factors is left for a later book and further empirical research.

Empirical studies which are available have also been almost exclusively confined to the term structure of interest rates. The major work has been under the auspices of the National Bureau of Economic Research and has involved research in the Corporate Bond market. Largely, this work has been of an historical nature to establish a basic yield curve representing, theoretically, the interest rate structure for the most risk free asset.[2] Utilizing these curves, studies of the Corporate Bond market were made in order to compare the "market rating" of bond quality to that of investment service ratings. The basic yield ("market rating") was found to vary more with the cycle than did agency ratings indicating that investors demanded and received a premium in excess of that justified by the actual risk of default.[3] A thorough analysis of the causative factors behind the differentials among various bond categories was deemed beyond the scope of these studies. The basic yield series for Corporate Bonds has been extended for the period 1952-1957 coupled with an analysis of the need for a thoroughgoing study of the entire structure of terms on which credit is extended and credit instruments transferred.[4] The only study concerned directly with yield differentials among monetary assets was of risk premiums on Corporate Bonds utilizing cross section analysis of

---

1. Conard, pp. 350-51.

2. D. Durand, "Basic Yields of Corporate Bonds 1900-1942," Technical Paper No. 3 (National Bureau of Economic Research, June 1942); D. Durand and W. Winn, "Basic Yields of Bonds 1926-1947: Their Measurement and Pattern," Technical Paper No. 6 (National Bureau of Economic Research, December 1947).

3. W. Hickman, Corporate Bond Quality and Investor Experience (New York; National Bureau of Economic Research, 1958); W. Hickman, The Volume of Corporate Bond Financing since 1900 (New York: National Bureau of Economic Research, 1953).

4. D. Durand, "A Quarterly Series of Corporate Basic Yields 1952-1957 and Some Attendant Reservations," Journal of Finance (September 1958), pp. 348-56.

a sample of firms. The basic conclusion was that risk of default and marketability are significant contributors to the average yield on particular bonds.[1]

Although the above review indicates that both theoretical discussion and empirical research have been confined almost exclusively to the consideration of the term structure of interest rates and have often imposed conditions involving risk free assets and perfect certainty, many of the theoretical implications drawn when these conditions are relaxed have bearing upon the interrelationships among yield curves in particular markets. Explanations of yield differentials among monetary assets fall under two general headings: (1) "pure" interest and (2) risk premium. The former breaks down into the consideration of transactions costs and the pattern of time preferences among individual investors. Under conditions of certainty or under conditions which allow close estimation of statistical probability based upon abundant empirical evidence, e.g., mortality tables, an investor may prefer a long term security for two basic reasons: (1) transactions costs are reduced as only one security is purchased to be held for the entire time period involved and (2) a fixed, periodic return is received over the entire time period. (For the sake of clarity, let us assume that the average, certain return on the short term securities equals the return on the single long term security.) Typically, however, investors are pictured as having a strong time preference and demanding a premium for extending the average maturity of their asset holdings. In a world of riskless securities and perfect knowledge, this typical explanation does not seem rational. Only when the second explanatory element is introduced involving the subjective estimation of the probabilities of default and of changes in capital values under conditions of uncertainty, can a consistent theory be approached. Portfolio selection involves the weighing of one element against the other as lower transactions costs and stability of income are evaluated against the probabilities that interest rates may rise or fall, plus the probability of default. It is this latter element, involving expectations and uncertainty, which is sorely in need of quantitative research as a first step in constructing an adequate theory of the entire interest rate structure.

It is the purpose of this study to isolate and analyze the determinants of the risk premium. To isolate the risk premium, analysis is confined to the Corporate and Federal Government Bond markets in which securities are readily substitutable and transactions costs are approximately equivalent throughout. Since bonds are classified by investment agencies according to default risk, it should be possible to construct yield curves for various homogeneous categories from the most risk free, or highest quality, to the lowest quality. By calculating the differential at specific maturities between the highest and lesser quality curves, it should then be possible to eliminate the last vestige of "pure" interest leaving only the existing risk premium.

That a risk premium must exist is evident from the fact that it is a function of the subjective estimation of the probability of (a) default with

1. L. Fisher, "Determinants of Risk Premiums on Corporate Bonds," Journal of Political Economy (June 1959), pp. 217-37.

respect to interest and/or principal and (b) changes in capital values. Government Bonds, to all intents and purposes, enjoy zero default risk. All other bonds, even if the situation is reduced to one determined investor who contends that a chance remains that, e.g., A T & T may default, must yield slightly more. The presence of a risk premium may thus be explained by "a" above. Before continuing, another factor should be noted. Changes in capital values arise primarily from changes in the level of interest rates. In a period of unfavorable expectations when maximum safety is desired, interest rates on the whole would tend to fall (prices rise) as investors adjust the risk content of their portfolios. Among bonds of varying quality, however, the probability of default risk would increase more at the lower end of the quality spectrum, tending to set up an opposing force to the over-all drop in yields (rise in prices). It might be predicted, therefore, that under varying conditions, yields (prices) of lower quality bonds will exhibit a relative degree of stability when compared with those of higher quality bonds.

Once appropriate yield differentials have been compiled for use as dependent variables, the hypothesis to be tested is that the risk premium between closely substitutable assets is a function of two basic elements (1) the relative supplies of such assets and (2) the demand for such assets—demand based upon expectations concerning conditions in the financial markets and the economy in general. Among the primary questions to be answered with respect to the first element is whether stocks and flows of assets at specific maturities are significant determinants of the yield differentials. If so, yield curves will seldom show a smooth curvature; if not, it is logical to conclude that a smooth function based upon variables common to the entire maturity range is an adequate approximation. Other supply questions to be answered concern the impact of stocks and flows in general, covering bonds, competing assets, new issues, etc. Given the stock of bonds outstanding, the question becomes what changes in the risk premium result from changes in the money market, the equity market and general economic conditions. The money market is the most closely attuned to changes in the interest rate structure; the equity market to changes in investor psychology and the relative attractiveness of equity versus debt; and existing economic conditions give a somewhat objective indication of default risk as corporate earnings respond to changes in business activity. Changes in any one of these areas affect expectations and also the availability of funds for the purchase of bonds. It is necessary to determine in what respect, if any, the analysis can support the contention that expectations must serve as the core of an adequate theory of the structure of interest rates.

2. The Bond Market

Although the bond market offers a close approximation of the theoretical model desired, it is far from perfect and is extremely deficient in meeting the three basic criteria of an acceptable financial market—depth, breadth and resiliency. It is a Walrasian-Hicksian affair.[1] Each weekday traders

1. See Hicks, chap. 9.

enter the market and cry bid prices (to buy) and ask prices (to sell). Sometimes only one trader appears in the market place and perhaps only a bid price is recorded that day; often no sale is consummated during the contracting period, and the traders retire until the next day. If failure to reach an equilibrium price persists for too long a period, the professional dealer enters the market and cries a price with the intention of adjusting the spread between bid and ask to a point which, in his opinion, more properly reflects existing market conditions. Finally, since the evaluation of risk is a highly subjective process based upon expected future conditions, actions of the monetary authorities and of the Treasury Department may temporarily influence prices especially during periods of exceptional uncertainty.

The above may lead to opinions that it is market imperfections which, in many cases, are the major contributory factor to the determination of yield differentials. However, with the degree of substitutability existing in the bond market and, therefore, the relatively favorable opportunities for arbitrage transactions, the primary effect of imperfections is merely to slow adjustments thus allowing one to minimize their impact by taking frequent observations over a period of time. It must again be noted that the size of the risk premium—especially that component which depends upon chances of capital gain or loss—is a function of subjective factors. By concentrating upon bonds, expectations concerning the general price level can be ignored, since monetary assets are all equally affected thereby; but expectations concerning the course of financial markets and general economic conditions are directly pertinent, and the normal diversity of opinion concerning them makes patent the high degree of uncertainty in any attempts to predict the future. Although taking observations over time should tend to reduce any impact due to imperfections in the market, this factor plus the fact that risk is strongly dependent upon highly diverse subjective evaluation means that any estimated equilibrium values of yield differentials can best be described in terms of a tendency or "normal" range.

3. The Differential

In order to calculate yield differentials—the dependent variable required for the purpose of statistical analysis—which most closely approximated the conditions outlined above, the first step was to choose bond categories which were homogeneous with respect to risk. The next step was to collect price data from which yield to maturity could be calculated and to plot these data against maturity for each bond category for a suitable time period. By fitting curves to the resultant scatters, it was then possible to construct a set of yield curves from which to calculate differentials for specific maturities.

The first category chosen was marketable U.S. Treasury Bonds. In the United States the prime securities consist of Federal Government obligations with the full taxing power of the central government behind them. The quality of these bonds is based upon the fact that they are, as nearly as possible, completely free of default risk. Yield curves for these bonds were

used as a base from which to calculate the dependent variables. Previous studies of bond yield differentials [1] have established a basic yield-maturity curve utilizing a visual fit of the lowest yielding (highest quality) bonds for the first quarter of each year 1900-1947. The primary purpose in constructing basic yield curves as opposed to relying upon Treasury Bond yield curves was to allow historical research during a time when there were variations in the tax status of Government Bonds and a shift in listing from the New York Stock Exchange to the Over-the-Counter Market. The present analysis covers the period 1954-1959 during which the status of Government Bonds was unchanged. It was therefore unnecessary to look further than this category for a basic yield. In order to get a sample of Corporate Bonds which were homogeneous with respect to risk, the specific categories chosen were Moody's Aaa and Baa bonds listed on the New York Stock Exchange. Moody's ratings are based upon the quality of Government securities and are the most widely consulted ratings by those who actively follow the bond market. Barring unforeseen shifts in industry solvency, such as occurred with respect to railroads after 1929, Moody's ratings have proved a stable and accurate indicator of the differences in default risk. [2] Aaa is the top quality rating while Baa is the minimum quality rating for investment purposes. The "Key to Moody's Ratings" defines each category as follows: [3]

Aaa

> Bonds which are rated Aaa are judged to be of the best quality. They carry the smallest degree of investment risk and are generally referred to as 'gilt edge.' Interest payments are protected by a large or by an exceptionally stable margin and principal is secure. While the various protective elements are likely to change, such changes as can be visualized are most unlikely to impair the fundamentally strong position of such issues.

Aa

> Bonds which are rated Aa are judged to be of high quality by all standards. Together with the Aaa group they comprise what are generally known as high grade bonds. They are rated lower than the best bonds because margins of protection may not be as large as in

1. See D. Durand, "Basic Yields of Corporate Bonds 1900-1942"; D. Durand and W. Winn, "Basic Yields of Bonds 1926-1947: Their Measurement and Pattern"; and D. Durand, "A Quarterly Series of Corporate Basic Yields 1952-1957 and Some Attendant Reservations."

2. See D. Durand and W. Winn, "Basic Yields of Bonds 1926-1947: Their Measurement and Pattern."

3. Moody's Industrial Manual (New York: Moody's Investors Service, 1960), pp. v-vi.

Aaa securities or fluctuation of protective elements may be of greater amplitude or there may be other elements present which make the long term risks appear somewhat larger than in Aaa securities.

A

Bonds which are rated A possess many favorable investment attributes and are to be considered as higher medium grade obligations. Factors giving security to principal and interest are considered adequate but elements may be present which suggest a susceptibility to impairment sometime in the future.

Baa

Bonds which are rated Baa are considered lower medium grade obligations, i.e., they are neither highly protected nor poorly secured. Interest payments and principal security appear adequate for the present but certain protective elements may be lacking or may be characteristically unreliable over any great length of time. Such bonds lack outstanding investment characteristics, and in fact have speculative characteristics as well.

For the three categories—Government, Aaa Corporate and Baa Corporate Bonds—monthly price data were collected over the six year period 1954-1959 inclusive.[1] This post war period was chosen as the longest feasible one during which both high and low interest rates prevailed and which was relatively free of changes in tax laws. Also, following the Treasury-Federal Reserve accord in 1951, the market was no longer subject to the influence of "pegged" rates for Government Bonds. Given these data, in theory one should find that two bonds with different coupons but the same quality rating and maturity have the same realized yield. In an attempt to assure at least an approximation of this result, only "straight" Corporate Bonds listed on the New York Stock Exchange, i.e., bonds not convertible into common stock or providing other special features such as warrants or voting rights, were included in the sample. However, a high degree of heterogeneity was found to exist in each category due to the appearance of new issues, the disappearance of maturing or called issues and changes in quality ratings. It was, therefore, necessary to include in the sample only those bonds outstanding in the pertinent category during the entire period under study. The result was monthly yield data for the six year period 1954 through 1959 for (1) 10-14 Government Bonds (2) 32 Aaa Corporate Bonds and (3) 27-31 Baa Corporate Bonds. Prices, from which yields to maturity were calculated, were taken as nearly as possible at the end of each month; where no actual sale information was available the midpoint between the bid and ask price was used. One further restriction was that during each annual

1. All price and volume outstanding data were collected from one of the following sources: Treasury Bulletin; Moody's Bond Record (New York: Moody's Investors Service); The Commercial and Financial Chronicle.

period consideration was given only to those bonds maturing in six years or more; this accounts for the variation in sample size.

With reference to Treasury Bonds except for two bonds in the first two maturity years, the sample consists of bonds subject to call at a specified date prior to maturity. (There were no noncallable bonds at maturities greater than seven years in 1954.) The Treasury Department follows the convention of figuring yield to call date, rather than maturity date, whenever a callable bond sells above par. Although this condition was quite prevalent in 1954 and existed for one long term bond much of the following two years, it was felt that following the convention would not be feasible in a study dealing with specific maturities. Any error involved seems minimal since, in 52 instances of a total of 144 cases during 1954 when the convention would have applied, the average adjustment to yield would be minus 3.3 basis points (.033 percentage points or an approximate error of 1.5%) with a range of from 0 to 8 basis points. In 20 instances occurring in the case of a single long term bond during 1955 and 1956 the average adjustment would be minus 3.7 basis points (a 1.2% error). Maturities on the bonds involved ran from 13 to 29 years with the call date five years in advance of final maturity. It seems dubious to suppose that an investor seeking a yield for such a lengthy period would count upon market conditions remaining the same and, therefore, immediately adjust his anticipated yield should the price exceed 100. Corporate Bonds are also, for the most part, subject to call, but the convention is not applied since the specific date is usually more nebulous than in the case of the Treasury Department. All yields, therefore, have been figured to final maturity.

The yield data, calculated as described above, were then plotted against maturity for the three bond categories for each of seventy-two months. The maturity for each bond was adjusted annually. For Government and Aaa Bonds yield curves were visually fitted to the resultant scatters; visual fitting was required since there is no accepted mathematical function to express the relation between yield and maturity and any attempt to obtain a smooth over-all curve would obviously dampen the sensitivity of the yield differential changes which are the subject of this study.

For the lower quality Corporate Bonds, however, a high degree of heterogeneity was found to exist within Moody's Baa classification. The fact that these bonds were less stable with respect to evaluation than Aaa's was anticipated, but the degree of their instability was not. A perusal of the "Key to Moody's Ratings" provides a clue as to why ". . . because of their very nature, rating changes are to be expected more frequently among bonds of lower ratings than among bonds of higher ratings . . ."[1] If agency rating changes were the only variations which occurred, there would be no problem since only bonds which have retained the same rating for the entire six year study period have been used as a source for the basic data. However, the market also "rates" bonds in the sense that price and, therefore, yield reflect changes

1. Moody's Industrial Manual, p.V.

in investor attitudes toward particular securities. "As might be expected, the variables emphasized by the market in its rating system are quite similar to the ones stressed by the investment agencies and the compilers of legal lists . . . The market . . . typically assigned top positions to the large, well secured, actively traded issues."[1] Hickman defined the market rating in much the same terms as the differential has been defined in this study—as the difference between the yield at market price and an equivalent point upon the National Bureau Basic Yield Curve.[2] He found a great deal of instability in this rating. "The market rating usually reflects changes in the credit standing of obligors more promptly than other ratings do . . . . The reason . . . is the extreme sensitivity, amounting almost to instability, of the market rating to changing conditions . . . "[3] Both agency and market ratings, then, are based primarily upon (1) the extent to which earnings cover interest and other fixed charges and (2) the variability of those earnings—the market rating being more responsive to actual, or anticipated, changes. By definition, earnings variability is more prevalent among the lower rated bonds adding to the risk of investing in a bond already close to the margin of safety with respect to debt coverage. Therefore, it is to be expected that the investor will consider more carefully the current, ever changing, characteristics of a Corporate Bond in the Baa category.

Upon the hypothesis that there might be some consistent individual bond yield biases over the period 1954–1959 due to variations in coverage of fixed charges and earnings among Baa Bonds, the following operation was performed. A second degree regression of yield versus maturity was estimated; the form of the equation was:

$$Y = a + b_1X_1 + b_2X_1^2 + c_1Z_1 + c_2Z_2 + \ldots + c_{11}Z_{11} \cdot$$

The dependent variable was the mean yield for an individual bond during each of the twelve semi-annual periods. X_1 equaled the maturity of the bond; any bias due to maturity was minimized since years to maturity was changed annually for each bond. Z_1 was a dummy variable equal to one for the semi-annual period January through June 1959 and zero for every other period; Z_2 was a dummy variable equal to one for the semi-annual period July through December 1958 and zero for every other period; etc. The results were significant (.99P) and produced a family of parallel curves with different intercepts for each six month period. The constant term indicates the intercept for the last six months of 1959; the "c" coeeficients indicate the factor to be subtracted from the constant term to derive the intercept for each of the semi-annual periods. The results of the regression and the calculated intercepts

1. Hickman, Corporate Bond Quality and Investor Experience, p. 301.

2. See Durand; and Durand and Winn. In this study the Treasury yield curve has been used as a basic yield curve as explained above.

3. Hickman, Corporate Bond Quality and Investor Experience, p. 18.

TABLE 1

B—Bond Adjustment
Regression Results

$$Y = 5.0258 + .027769\,X_1 - .000503\,X_1^2$$

(.13054) (.010429) (.000210)

COEFFICIENTS	STANDARD ERROR	CALCULATED INTERCEPTS
$c_1 = -.25629$	(.11398)	4.77
$c_2 = -.32830$	(.11296)	4.70
$c_3 = -.40830$	(.11296)	4.62
$c_4 = -.22376$	(.11201)	4.80
$c_5 = -.70755$	(.11201)	4.32
$c_6 = -1.0208$	(.11207)	4.01
$c_7 = -1.5122$	(.11207)	3.51
$c_8 = -1.6124$	(.11120)	3.41
$c_9 = -1.7008$	(.11120)	3.33
$c_{10} = -1.7586$	(.11038)	3.27
$c_{11} = -1.6679$	(.11038)	3.36

$R^2 = .717$

$Su = .041880$

are shown in Table 1. To check whether the slope of the over-all regression was consistent for every six month period, an F-test was made utilizing the sums of squared residuals from independently run semi-annual regressions and from the over-all regression. The results (F = .6693 with 22/312 degrees of freedom) supported the null hypothesis that there was no significant difference in the slope of the yield curve for different periods. Adjustment factors were then computed as the difference between Y_C and Y_O, where Y_C equals the individual mean computed from the regression equation and Y_O equals the actual six month mean yield for an individual bond. (See Appendix.) The adjustment factors so calculated failed to exhibit the consistency necessary to eliminate the wide dispersion among the Baa Bonds. In some cases the "market" rating of an individual bond deteriorated, changing from plus to minus between 1954 and 1959; in others it improved, changing from minus to plus.

Investor confidence was not constant during the six years, and it is ultimately upon this factor that the market rating depends. "The market rating provides a sensitive yardstick for ranking issues in order of default risk at any given moment; but the yardstick is elastic expanding and contracting with investor confidence."[1] 1954-55 was a period of boom and steady improvement; 1956-57 was a period of leveling off; and 1958-59 was a period of recession and uncertainty. At the same time earnings experience for individual corporations varied. Baa bonds are largely comprised of railway bonds, and it is therefore proper to use a small sample of these as an illustration of changes in relative standings among individual bonds. Comparing three Class I (large) Railroads in the sample, numbers 29, 24 and 12 in the Appendix, against all Class I Railroads, the pattern is as follows:

Operating Record—Percentage of Gross Revenue Carried through to Net Operating Income before Federal Income Taxes[2]

	Class I Roads	#29	#24	#12
1954	11.7	21.1	9.2	11.0
1955	15.3	26.1	10.0	15.2
1956	13.9	23.5	8.1	11.8
1957	11.9	20.2	9.0	9.3
1958	10.5	18.5	9.3	4.0
1959	10.4	22.8	10.4	4.8

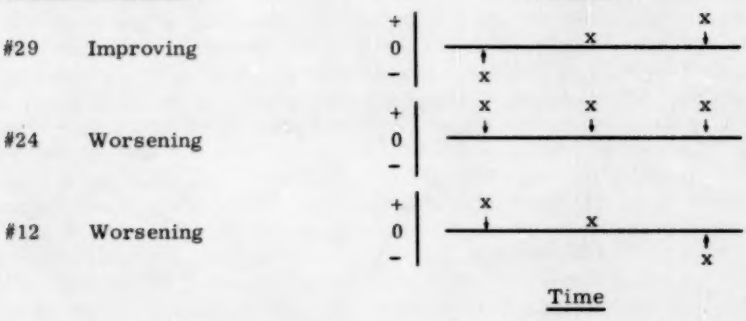
Debt coverage for the three roads showed changes of:

#29	+ 30%
#24	- 7%
#12	- 52%

1. Ibid, p. 355.

2. Moody's Transportation Manual (New York: Moody's Investors Service, 1960), p. a40.

Deviations from the regression ($Y_C - Y_O$) were of the following general pattern:



It was this type of variation in relative market rating within the general Baa classification which had to be removed in order to consider changes in yield differentials ascribable to the quality rating as a whole. A further random sampling of the Baa bonds confirmed the fact that relative standing is a function primarily of earnings and debt coverage, and also that these market ratings are very responsive to changing aspects of the corporate balance sheet.

It was evident from the pattern of residuals in the Appendix that the six year period of the study could be roughly divided into three two year periods—1954-55, 1956-57, and 1958-59—and an average adjustment applied for each of the three periods. In almost every case there was a change in relative position during 1956-57 as the economy leveled off and again in 1958-59 as the economy turned down. In 1954-55 and 1958-59 the market ratings showed some degree of stability and in every case, except five in the latter period, the adjustment factor shown for these periods is the same for the four six month periods encompassed. The 1956-57 period, however, does not exhibit the same degree of stability as re-evaluations took place at different times based upon the earnings trends of specific corporations. It was necessary, therefore, to estimate subjectively the cut off point for a particular market rating. In the case of bond #5, for example, it was evident that a drastic re-evaluation took place in mid-1956 (see Appendix). Inclusion of the first half of 1956 with the next three semi-annual periods would have called for a mean adjustment of -62 basis points during 1956-57. The result would have been a vast overstatement of the market rating in the first half of 1956 and an understatement of it thereafter. It was deemed more appropriate to extend the initial division through the first half of 1956. In the case of bond #9 the market rating did not change until 1957; in #10 there was a drastic re-evaluation in the last half of 1957; #20 lagged twice, not adjusting until the second half of 1956 and again in the second half of 1958; etc. The final Baa bond adjustment, therefore, was calculated predominantly as

Chart 1*

a.

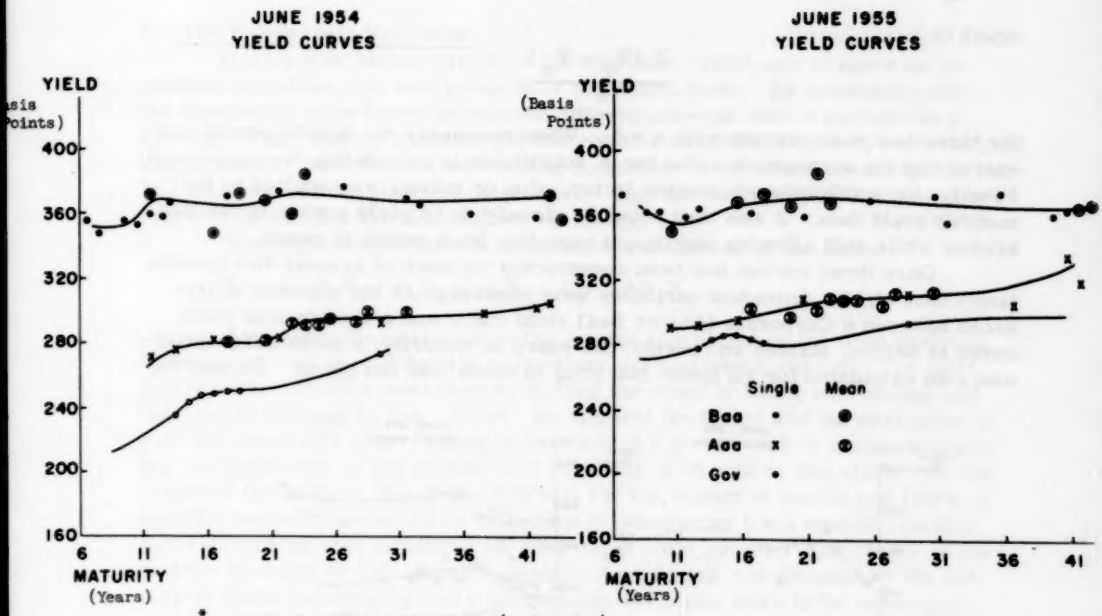
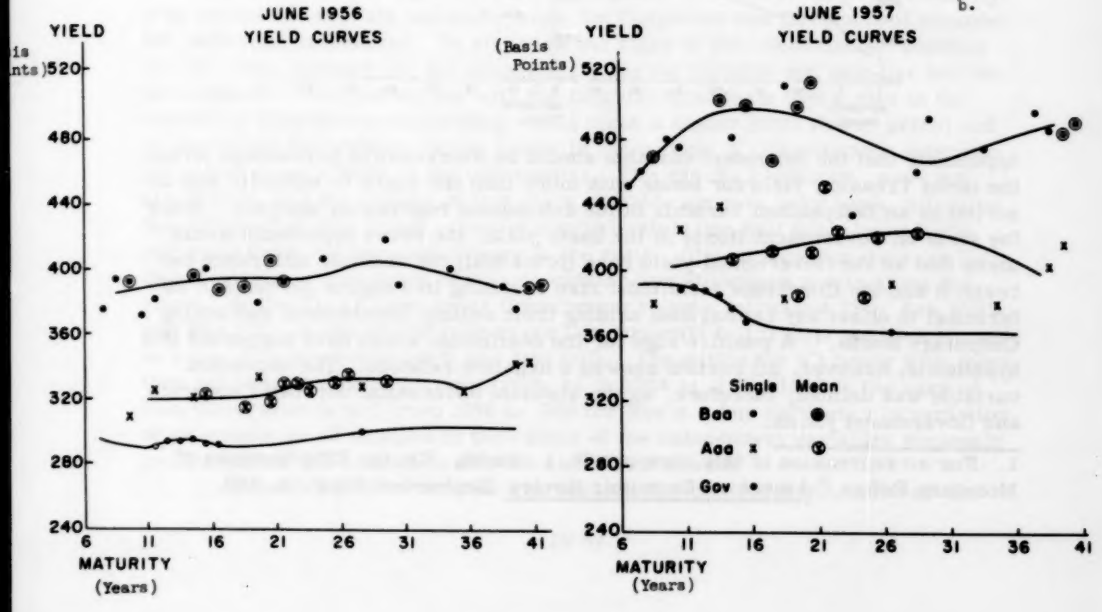


*Chart No. 1 consists of 3 pages (a, b and c).

Chart 1

b.

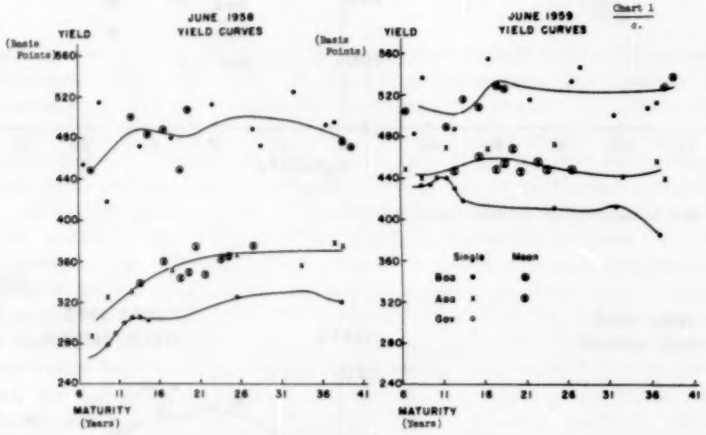


equal to

$$\frac{\Sigma (Y_c - Y_o)}{n}$$

for three two year periods with n = 4 . When necessary the middle period was varied and the appropriate value for n substituted in calculating the adjustment. Finally, the applicable adjustment factor, plus or minus, was applied to the monthly yield data. It was then possible visually to fit yield curves to the Baa scatter while still allowing significant variation from month to month.

Once three curves had been constructed for each of seventy-two months (see Chart 1), the dependent variables were measured as the absolute differences between a Corporate (Aaa or Baa) yield curve and a Government yield curve at twelve, sixteen and twenty-one years to maturity; a mean differential was also calculated for all bonds maturing in more than ten years. To test the



hypothesis that the dependent variable should be expressed in percentage terms, the mean Treasury yield for bonds with more than ten years to maturity was inserted as an independent variable in the subsequent regression analysis. Since the yield on Government Bonds is the basic yield, the above hypothesis would mean that as the Government yield rose (price fell) the absolute difference between it and the Corporate yield must rise resulting in a higher percentage differential to offset any capital loss arising from selling Government and buying Corporate Bonds.[1] A positive sign for the coefficient would have supported this hypothesis; however, all results showed a negative relation. The dependent variable was defined, therefore, as the absolute differential between Corporate and Government yields.

1. For an expression of this view see W. L. Smith, "On the Effectiveness of Monetary Policy," American Economic Review (September 1956), p. 590.

4. The Independent Variables

Having now defined the yield differentials, which are to serve as dependent variables, the next problem is to explain them. As mentioned above, the hypothesis to be tested is that the differential—the risk premium—is a function of two basic elements: (1) the relative supplies of closely substitutable assets and (2) expectations concerning conditions in the money market, the capital markets and the economy in general. This section presents the procedure followed in defining the independent variables.

The regression analysis undertaken was linear, of the form:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 .$$

Before discussing the variables which proved significant in the final analysis, however, it may be worthwhile to consider several which did not. The first independent variables considered involved the stock of bonds outstanding and changes in this stock, i.e., flows. Investment decisions and the evaluation of risk are made at a point in time in the face of a given stock of bonds outstanding and knowledge of the present and past rate of change in this stock. At the outset of the study it was desired to test for the impact of stocks and flows at specific maturity points as an indication of whether or not a smooth function can serve as an approximation for long term yield curves. The choice of maturities at which to calculate the dependent variables was dictated by the pattern of bonds outstanding since independent variables were to be used equal to the ratio, and the change in ratio, of volume outstanding (Corporate/Government) at specific maturities. (See Table 2.) Ratios were tested for (1) a three year range of outstandings with the maturity year of the differential as the midpoint (2) a four year range and (3) for all bonds over ten years from maturity. Also tested as separate variables were the Corporate and Government volumes for particular maturities. In all cases the signs of the coefficients—positive for the ratio, positive for the separate Corporate variable and negative for the Government—were consistent with the original hypothesis that a rise in the volume of Corporates outstanding would mean a higher yield (lower price) and an increased differential, while a rise in the volume of Governments outstanding would cause a drop in the differential. The results, however, were not significant, and the impact of this stock variable appears very weak, affecting the differential only to the extent of approximately one half basis point (.005 per cent) for a one per cent change in the ratio. The explanation does not lie in any degree of constancy of volume outstanding at various points along the curve. In the 11-15 year range Government volume increased 32%, Aaa 100% and Baa 10%; in the 16-20 year range Governments fell from a positive amount to zero, Aaa increased 100% and Baa 60%. The ratios for all bonds with more than ten years to maturity rose from .11 in 1954 to .20 in 1959 in the case of Aaa/Governments and from .056 to .068 for Baa's. Any deficiency in variation of or magnitude of changes in the values of the independent variables obviously cannot explain the failure to achieve significant results.

Initially the flow variable was derived directly from the above as the

TABLE 2

Distribution of Volume Outstanding

1954–1959

Maturity in Years	Aaa						Gov.						Baa					
	54	55	56	57	58	59	54	55	56	57	58	59	54	55	56	57	58	59
10	———					00	———			00	00	00	00	———				
11	———						———				00	00	——	00	———			
12	———						———					00	———	00	———			
13	00	———					———						———	00	———			
14	—0	00	———				00	———					———	0	———			
15	——	00	00	———			——	00	———				———					
16	———	00	00	———			00	——	00	———			———					
17	———		00	00	——		00	00	—0	00	———		———					
18	———			00	00		00	00	00	00	00	——	———					
19	———					00	00	00	00	00	00	00	———					
20	———						00	00	00	00	00	00	———					
21	———						——	00	00	00	00	00	———					
22	———						00	00	00	00	00	00	00	———0	———			
23	———						00	00	00	00	00	00	——	00	——	00	———	
24	———						——	00	00	00	00	00	———	00	——	00	——	
25	———						00	——	00	00	00	00	———		00	——	00	
26	———						——	00	——	00	00	00	———			00	——	
27	———						00	——	00	——	00	00	———0	——	00			
28	———						00	00	00	00	——	00	00	———	0———			
29	———						00	00	00	00	00	——	——	00	———			
30	———						00	00	00	00	00	00	0———	00	———			

——— indicates bond(s) outstanding for period designated.

00 indicates no bond outstanding for period designated.

rate of change in the stock variable. In this instance the constancy of the variable was marked as most changes of any consequence were due to maturity shifts; new issues were confined almost exclusively to intermediate (5-10 years) and very long term (25-35 years) maturities. Results were not significant. To make a further check of the hypothesis that flows are important determinants of the yield differential, independent variables were tried representing (1) the sum of new issues—public and private and (2) the ratio of Corporate new issues to Governments.[1] Again results were not significant. The explanation probably lies not in the fact that new issues are unimportant but rather in the fact that new issues are floated with utmost care in order that they may be fitted into the existing interest rate structure with a minimum of upset. In the private sector investment bankers support new issues, i.e., support the price at the initial level until the new bonds are absorbed. If the bonds are not absorbed, the issue may even be withdrawn. In the government sector the Treasury Department chooses its time carefully. Except for one small cash issue in July 1955 when books were reopened for a 39 year 7 month bond, cash financing during the period under study was limited to maturities under five years until 1957. Only during the recessionary period from September 1957 to June 1958 when speculative demand for long term fixed interest securities ran high did the Treasury venture into the long term capital market. Even then when the market threatened to become disorderly following the June 1958 issue, the Federal Reserve and the Treasury stepped in to support the prices of Government Bonds. Statistical measure of the impact of these issues was made impossible by their random nature and also by the fact that psychological factors due to the high degree of uncertainty accompanying any decline in the economy overcame any consistent relation which might have existed.[2]

In short it was not possible to support the hypothesis that the stock at specific maturities along the yield curve nor the flow variables have any important impact on the differential. One is able to conclude from this that the interest rate structure for long term bonds can be approximated by a series of smooth curves without the necessity of allowing for humps and/or discontinuities at various maturity points due to distinctly different proportions among outstandings. The long term bond market is relatively homogeneous (in this respect) subject only to general factors common to the over-all maturity range. The final equation thus includes two variables dealing only with gross volume and three variables bearing upon financial markets and the state of the economy. They are defined as follows:

1. The ratio of the total par value of Corporate Bonds listed on the New York Stock Exchange to the value of competing long term debt. Competing long term debt was defined as the sum of equity stocks outstanding plus mortgage debt outstanding. The volume of equities was computed

1. Source: Federal Reserve Bulletin.

2. For a similar conclusion based upon empirical analysis see Conard, p. 335 and especially, p. 338 (top).

by dividing Dividends, as published under Personal Income and Disposition of Income in the Department of Commerce's National Income Accounts, by Moody's Stock Yields. The volume of municipal bonds was the total par value as published in the Annual Report of the Secretary of the Treasury. For the last six months of 1959 it was necessary to estimate the figures through a conversion ratio utilizing Bond Buyer data based upon offer prices as published in the Survey of Current Business. Mortgage debt outstanding was taken from the quarterly figures published in the Federal Reserve Bulletin and the difference prorated over each quarter in order to have a monthly figure.

2. The ratio of the total par value of Government Bonds (as opposed to notes, certificates or bills), as published in the Treasury Bulletin, to the value of competing long term debt as defined above.
3. The interest rate on prime commercial paper (4-6 months) as published by the Federal Reserve Board.
4. The dividend yield on equities as measured by Moody's Stock Yields covering a sample of 200 stocks representing a cross section of industry.
5. The seasonally adjusted rate of unemployment in the civilian labor force as published now by the Bureau of Labor Statistics and formerly by the Bureau of Census. Prior to 1957 it was necessary to adjust the published data in accordance with the new seasonal adjustment factors in current use and to remove from total employed those on temporary layoff or waiting to start a new job or business. For 1954 the latter adjustment was estimated but should be very close to the actual figures (which were not available in published form at the time these computations were undertaken).

5. Results

Table 3 and Chart 2 show the regression results for seventy-two monthly observations 1954 through 1959. Column one of Table 3 indicates the particular equation with, for example, A designating the dependent variable as Aaa Corporate minus Government yield and the number designating the specific maturity year at which the differential was measured: A > 10 indicates that the dependent variable was derived from the difference in mean yields for all Aaa and Government Bonds with more than ten years to maturity. Similarly B designates equations involving Baa Corporate Bonds. Values in the table are defined as follows:

a = the constant term

b_i = the slope of the ith independent variable; i = 1, . . . 5, in accordance with the above listing of independent variables

s_{b_i} = the estimated error of b_i and is indicated in parentheses below each coefficient; an asterisk indicates Student's t-Test (b_i/s_{b_i}) for testing b_i significant at the .95P level

TABLE 3

Regression Results

	a	b_1	b_2	b_3	b_4	b_5	R^2	Su
A-12	-.24147	.08771	-.13825	-.06737*	.24670*	.00294	.491	.09381
	(.21606)	(.06847)	(.08278)	(.02754)	(.04573)	(.01964)		
A-16	-.17883	.12900*	-.21779*	-.04936*	.31698*	.02665	.694	.08209
	(.18907)	(.05991)	(.07244)	(.02410)	(.04002)	(.01718)		
A-21	-.30947	.17118*	-.25325*	-.02426	.26097*	.02064	.690	.07621
	(.17552)	(.05562)	(.06725)	(.02237)	(.03715)	(.01595)		
A>10	-.31802	-.00303	-.10342	-.11668*	.49525*	.02997	.819	.07396
	(.17034)	(.05398)	(.06725)	(.02171)	(.03606)	(.01548)		
B-12	-.96560*	.22676*	-.42824*	-.22675*	1.0491*	.12482*	.853	.14864
	(.30458)	(.11275)	(.13797)	(.04535)	(.07622)	(.03173)		
B-16	-.61794*	.3320*	-.55329*	-.14803*	.87254*	.11497*	.877	.12199
	(.24998)	(.09253)	(.11324)	(.03722)	(.06256)	(.02604)		
B-21	-.86398*	.32229*	-.51053*	-.06837	.75392*	.11963*	.834	.13585
	(.27837)	(.10305)	(.12610)	(.04145)	(.06966)	(.02900)		
B>10	-.23595	.14052	-.38585*	-.21284*	1.1611*	.13221*	.933	.10021
	(.20535)	(.07602)	(.09302)	(.03057)	(.05139)	(.02139)		

R^2 = the coefficient of multiple determination

s_u = an unbiased estimate of the standard error of residuals

F-Tests for each equation showed R^2 to be significant at the .99P level. To test the null hypothesis that there was no significant difference among equations A-12, A-16, and A-21, and similarly among B-12, B-16, and B-21, observations from each set of three equations were pooled and an F-Test made with:

$$F = \frac{\frac{\Sigma u'_j{}^2 - \Sigma\Sigma u_j{}^2}{n}}{\frac{\Sigma\Sigma u_j{}^2}{N}}$$

$\Sigma u'^2_j$ = sum of the squared deviations (residuals) from the pooled equation

$\Sigma\Sigma u^2_j$ = sum of the sum of the squared deviations from the individual equations

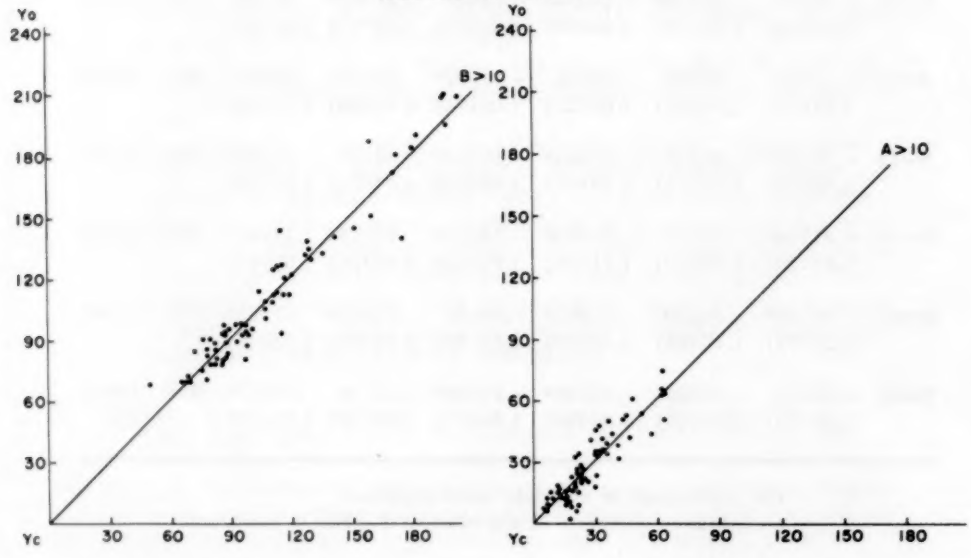
n = 10

N = 201

In both cases the F-Test indicated a significant difference at the .99P level among the individual equations: for the A equations F = 17.326; for the B equations F = 9.684; for .99P F = 2.41.

Chart 2

OBSERVED DIFFERENTIAL PLOTTED AGAINST COMPUTED DIFFERENTIAL



6. Analysis of the Results

A. Stock Variables

The first two independent variables, the ratios of Corporate and of Government Bonds to competing long term debt, are measures of the stock of bonds outstanding relative to major competitive instruments in the capital market. These variables serve as adjustment factors so that variables reflecting the more subjective elements in risk evaluation may be analyzed "given the relative volume of bonds outstanding." Between 1954 and 1959 the stock of both Corporate and Government Bonds fell relative to other assets. The values of the independent variables were:

	January 1954	December 1959
X_1	32.6	17.8
X_2	24.7	12.6

As the volume of Corporate Bonds fell relative to other assets, the original hypothesis was that prices would rise, yields fall and the differential decrease; as the Government ratio fell the hypothesis was that prices again would rise, yields fall but in this case the differential would increase. The signs of the coefficients, positive for the Corporate ratio and negative for the Government, supported the hypothesis. The values were significant, at least to the .90P level, except for equations A-12 and A>10; the Government ratio exhibited the stronger impact throughout. Further the impact of these variables tended to be weaker at 12 than at 21 years. In view of the fact that two of the three expectational coefficients (b_3 and b_4) exhibit the opposite tendency, it may be concluded that subjective evaluation of risk based upon expected short run changes in financial markets dominates the more objective influence of the stock variables as the long term bond market (over ten years) becomes more nearly contiguous to the intermediate market (five to ten years).

B. Prime Rate on Commercial Paper

The prime rate on commercial paper (4-6 months) was chosen as an indicator of expectations of changes in the money market and of the availability of short term funds. The rediscount rate was not used since it is much more "sticky." Also it was desired to have a rate at least one step removed from the direct institutional control of the Federal Reserve Board yet strongly subject to its actions. At the other end of the spectrum the prime rate on bank loans might have been used, but, again, changes are relatively infrequent and more the result of institutional decisions than the free play of the money markets. The commercial paper rate was used as being the most responsive to changes in liquidity and as the best indicator of the general trend in interest rates—basically as the determinant of that portion of the risk premium demanded to compensate for the probability of a change in capital values due to changes in the interest rate structure. As interest rates fall, bond prices rise, the reward for undertaking the risk of capital loss is reduced and one expects the differential to increase if investors are to be induced to maintain approximately the current, over-all portfolio composition. The negative signs of the coefficients support this hypothesis. Further, as maturity lengthens, the impact of the interest rate lessens; at 21 years the value is not significant. This is explained by the fact that the very long term market is more of a stable, investor's market not as subject to temporary fluctuations in the supply of funds nor as likely to attract speculators with short time horizons who are concerned with the possibility of capital gains from expected changes in interest rates.

For the A equations the interest rate was significant only when

introduced with a lead of one month, i.e., at time period t + 1. The high quality bond market is by far the most sophisticated market—largely peopled by institutional investors. These professional investors keep themselves extremely well informed and are in a better position than most to anticipate changes in the interest rate structure. Generally liquidity changes in the banking sector prior to changes in the commercial paper rate, and these investors are well aware of the fact. This contention is supported by the results of a simple regression of the form:

$$r = f(\text{net free reserves})_{t-1}$$

The significant equation which resulted was:

$$\begin{array}{ccc} Y = 2.9125 & - & .00216X \\ (.057) & & (.00014) \end{array}$$

The correlation coefficient, R, was equal to .88 and the computed change in the commercial paper rate (r) was approximately 20 basis points for every 100 million dollar change in net free reserves. This indicates considerable predictive value pertaining to the money market can be derived from a high degree of familiarity with the liquidity position of financial institutions. Substitution of net free reserves for the commercial paper rate as an independent variable in the equations showed some significant results but not the consistency of the variable adopted. For the B equations neither the commercial paper rate introduced with a lead nor the substitution of net free reserves proved significant—proving that in this less sophisticated market expectations are dominated by existing conditions. Sophisticated investors could easily enough invade the Baa market but part of sophistication is obtaining return while undertaking minimum risk. In perilous times high quality bonds offer the possibility of return with only limited risk; Baa Bonds offer a return but only with increasing risk of default as Corporate earnings drop. In favorable times the risk of holding alternative assets decreases, and the sophisticated investors, enscounced in high quality bonds, make the first move. In either case, it is the high quality market which sets the pace.

C. Equity Yield

The dividend yield on stocks is the best available measure of investor psychology—of expectations, uncertainty and the relative attractiveness of monetary versus real assets. Since it is not a perfect measure, especially with respect to the bond market, there is a degree of uncertainty unaccounted for; but it still serves as an approximation for the determinant of that portion of the risk premium which is a function of the probability of capital loss or gain due to relative price changes between equities and bonds. The dividend yield was chosen rather than the earnings yield in this case since dividends are a much more stable element than are earnings. Thus, the variable is primarily a function of buying and selling pressures in the stock market rather than a combination of this desired factor and a cyclical element introduced by

earnings fluctuations consonant with business conditions. With a buoyant stock market (falling yield), favorable expectations of business activity and a rising general price level, a preference for equity assets develops coupled with a willingness to expand the proportion of risk in portfolios. Between markets investors switch from bonds to stocks and bond prices fall; within the bond market the probability of capital gain increases and one expects the differential to decrease. The fact that the coefficients show a positive relationship bears out this hypothesis. It should also be noted that the values of the coefficients indicate that this variable is by far the most important of the expectational variables.

For the A equations no consistent trend in the coefficient values is evident. This high quality market is an institutional one dominated by those dealing primarily in bonds. For the B equations the coefficients show a tendency to decrease in value for the longer maturities indicating the pressure of offsetting evaluations of risk. Given a falling stock market (rising yield), investors seeking capital gain can reduce risk by switching from stocks to bonds. This action tends to bid Baa prices up, yields fall and the differential narrows. On the other hand investors seeking safety within the bond market can reduce risk by switching to higher quality bonds. This action tends to depress Baa prices, yields increase and the differential widens. In the very long term market it is evident that those seeking capital gains dominate, while the shorter term and Aaa markets are more strongly influenced by conservative investors who must balance liquidity against profits and risk.

D. Unemployment Rate

The percentage of the civilian labor force unemployed was designed to serve as a proxy for the subjective evaluation of default risk. It is a rough measure of the existing state of the economy upon which profits depend and ultimately debt coverage and default. As unemployment rises, profits normally fall, the risk of default increases, and one would expect the differential to increase. The positive signs of the coefficients support this hypothesis. The unemployment variable is basically weaker than those representing the financial markets. In addition, the coefficients proved not significant for the A equations. For both the A and the B equations the values of the coefficients were approximately equal for all points along the yield curves. This result might be expected. Aaa Bonds consist mainly of utilities plus some "blue chip" industrials and rails, e.g., General Motors, Standard Oil (N.J.), Texas Company, Atchison, Topeka and Santa Fe Railroad, etc. The possibility of default is remote barring another "Great Depression." Aaa is the top quality rating assigned by Moody applying as well to Government Bonds; the results indicate acceptance of the high degree of substitutability so indicated and further supports the view that, for a given stock, the most important functional relationship in determining bond yield differentials involves risk and expected changes in the capital values of monetary assets. In contrast the Baa bonds are mostly railroads, many of which have been experiencing a declining profit trend. The consistency of the values follows from

the fact that maturity does not grant one a prior claim on the assets of a bankrupt corporation and further bears out the plausibility of a general quality rating for categories of bonds over the long term.

7. Alternative Hypotheses

Upon plotting the residuals from the above equations against time, it was noted that the dispersion showed some tendency to increase during periods of business uncertainty. In an effort to explain this dispersion, it was decided to test the following hypotheses:

(1) the rate of unemployment is not an adequate indicator of the economy
(2) the increased scatter is attributable to a change in the function during periods of business recession.

The latter hypothesis would be supported if it could be shown that either:

(a) there was a change in the value of the coefficients, i.e., the slope; for example, reactions to the yield on equities might be greater in a recession than otherwise, or
(b) the constant term was significantly different; for example, the risk premium might be greater in a recession than otherwise.

The independent variable (X_6) chosen as a substitute for the rate of unemployment (X_5) was the deviation from the trend of the Industrial Production Index (1957 = 100).[1] A trend line was fitted to the seasonally adjusted monthly index for the period July 1953 through June 1960 and the positive and negative deviations ($Y_O - Y_C$) computed. This variable (X_6) was substituted for the rate of unemployment (X_5) in the regression. In addition, to test for changes in the slope, dummy variables of the form $d \cdot X_3$, $d \cdot X_4$ and $d \cdot X_6$ were introduced where $d = 0$ for a positive deviation and $d = 1$ for a negative deviation (X_3 and X_4 refer respectively to the interest rate and equity yield variables as heretofore defined). To test for changes in the constant term, a dummy variable, D_O, was introduced where $D_O = 0$ for a positive deviation and $D_O = 1$ for a negative deviation. The result was that of all the variables tested, separately and conjointly (including X_6), only $d \cdot X_6$, the product of the dummy and the computed deviation, was significant. The hypothesis that the constant term was significantly different under adverse business conditions as measured by the Industrial Production Index was not supported; the hypothesis that investor reactions to the interest rate or to equity yields differed significantly during recessions was not supported. The hypothesis that the rate of unemployment was an inadequate indicator of economic conditions was not supported, (a) since the variable representing deviations from the trend of the Industrial Production Index (X_6) did not exhibit a statistically significant relationship and (b) although the coefficient for the variable $d \cdot X_6$ was significant, when substituted for the rate of unemployment (X_5), it did not materially improve R^2, the standard error of estimat nor did it alter the trends in other coefficient values or the scatter of residuals.

1. Source: Federal Reserve Board, Industrial Production 1959 Revision (Washington, 1960).

These results indicate that economic conditions, as measured by the Industrial Production Index, are only significant when adverse and that under adverse economic conditions there is no significant difference in the other coefficients (b_3 and b_4). It was concluded that improvement was insufficient to justify changing the original equation.

In passing, another relationship which might have bearing on this study should be mentioned, i.e., the relationship between bond yield differentials and the future course of the stock market. Each Monday Barron's, a financial weekly, publishes as a Confidence Index the ratio of the yield index on high quality bonds to its over-all bond index. Recently this Confidence Index has been presented as an instrument for predicting the trend in the stock market with a lead time of from sixty to ninety days.[1] To test this assertion various lead times were introduced into the original equations both with respect to equity yields and the rate of unemployment. At no time were significant results obtained. Admittedly the dependent variable in this case is the absolute differential rather than a ratio, but it is dubious that this fact alone explains the lack of statistical significance in the above results. Certainly the high quality bond market is populated by sophisticated investors who often can and do anticipate changes in economic conditions and the financial markets, but there seems little indication that, in a world subject to great uncertainty concerning the future, bond yield differentials are other than coincident indicators when a consistent relationship over time is desired.

8. Expectations

As mentioned above, when the residuals from the regressions are plotted against time, an increased dispersion is evident during periods of unfavorable business conditions such as the 1957–58 recession (see Chart 3). Taking those deviations which exceed one standard error during 1957 and 1958, it is possible to discern a rough pattern (see Table 4). Since statistical analysis has left this pattern inviolate, it seems pertinent to assume that the "goodness of fit" depends primarily upon the elasticity of expectations with respect to the interest rate. Especially in periods of great uncertainty there occur sudden changes in expectations which in a market lacking depth, breadth and resiliency may cause wide divergence of observed from computed values before the differential can find its "normal" equilibrium. As Hicks says, "... the sort of variations in interest rates which are consistent with quiet times and with the maintenance of organized markets are quite small; for, as we have seen, the level of interest rates ultimately measures the intensity of a certain set of risk factors, and this intensity is unlikely to remain for long outside certain broad limits. Consequently, when the rate of interest (any rate of interest) rises or falls very far, there is a real presumption that it

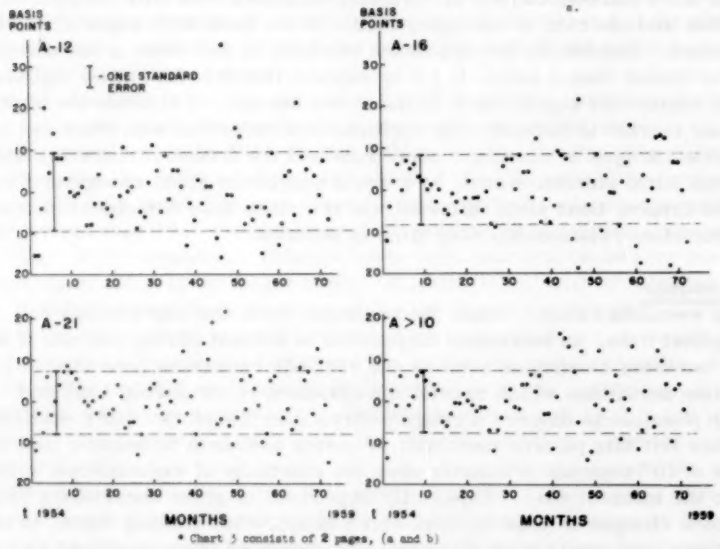
1. See J. E. Granville, "Market Forecaster?" Barron's (September 7, 1959), p. 9; also J. E. Granville, A Strategy of Daily Stock Market Timing for Maximum Profit (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960), pp. 101–22.

will come back to a 'normal' level. This consideration would seem to prevent interest-expectations from being very elastic."[1] Although Hicks refers to the "level" of interest rates, the context is such that his remarks are applicable to the entire interest rate structure and therefore to yield differentials. He goes on to say, in a footnote, the "short rate will only react quickly upon the long rate if there is immediately apparent some significant reason" and "this creates an expectation that high short rates are less likely in the future than they were in the past."[2] The implication is that during prosperous, basically stable times the structure of interest rates tends to be accepted as given, i.e., expectations are inelastic. Investors react to change

RESIDUALS PLOTTED AGAINST TIME

Chart 5*

a.



* Chart 5 consists of 2 pages, (a and b)

only when convinced the change is permanent. Only when good and sufficient reason becomes evident are changes interpreted as harbingers of the culminating point of a given trend, i.e., expectations take on a negative elasticity. The concept of negative elasticity of interest rate expectations must be approached with some trepidation. It is a situation which cannot last for more than one or two time periods since otherwise one would be in the position of arguing that, e.g., a continually falling interest rate would mean that the

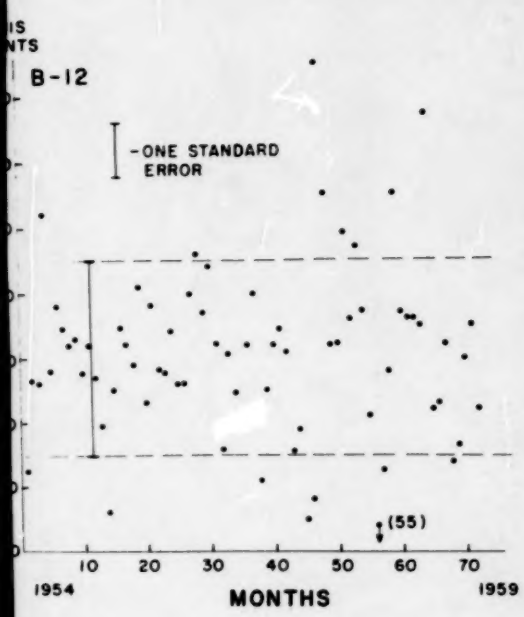
1. Hicks, p. 262.

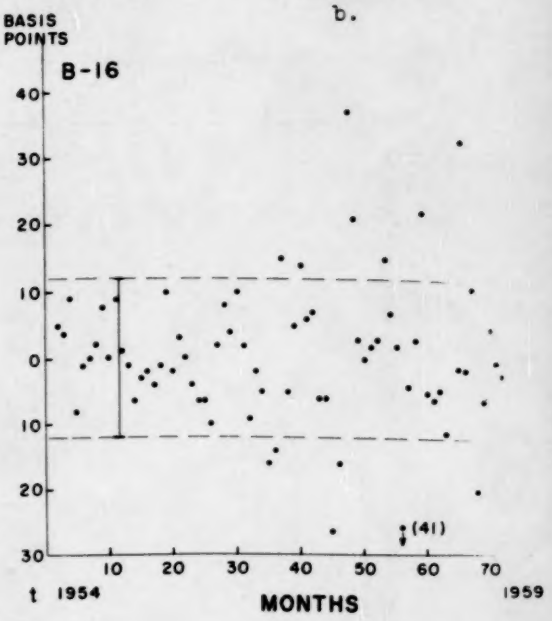
2. Ibid.

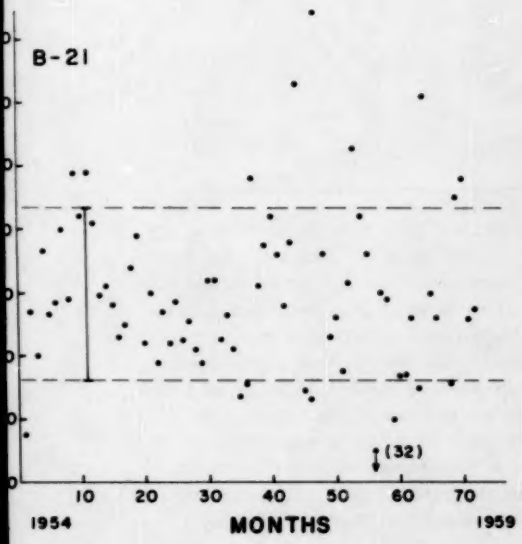
RESIDUALS PLOTTED AGAINST TIME

Chart 3

b.







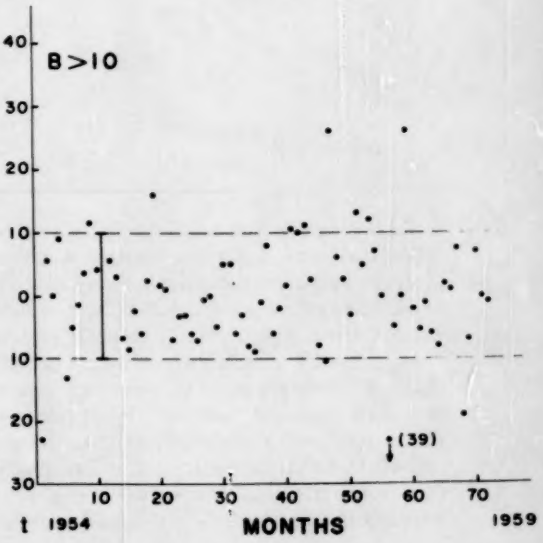


TABLE 4

Pattern of Deviations Exceeding One Standard Error

		A-12	A-16	A-21	A-10	B-10	B-12	B-16	B-21
1957	J		+	+				+	+
	F	−					−		
	M								
	A			−	−			+	
	M			−		+			
	J		+	+	+	+			
	J			+	+	+			
	A			+	+				
	S	−	−				−	−	−
	O	−	−			−	−	−	−
	N	+	+	+	+	+	+	+	+
	D				−		+	+	
1958	J			−	−				
	F	+							
	M	+			+	+	+		
	A		−						
	M					+	+	+	+
	J								
	J								
	A	−	−	−	−	−	−	−	−
	S			−			−		
	O								
	N		+		+	+	+	+	−
	D	−	−	−	−				

expected rate would be continually revised upward. This may be true in a cyclical swing as investors foresee a greater reaction back to "normal" the lower interest rates drop during a recession, but there are some limits as to the credibility of maintaining that such a position can exist in any long run sense. Expectations would more normally become elastic as a downtrend in rates is "recognized," tend to become only briefly negative as the turning point is anticipated and then tend to elasticity again as the turn is made and interest rates rise back to their "normal" level (at which inelastic expectations prevail). The more unexpected the changes are, the greater the elasticity of expectations as investors suddenly believe they can recognize a trend and, upon the basis of these expectations, revise their evaluations of the riskiness of competitive assets. Also, there is a decided degree of asymmetry in this process as

expectations tend to exhibit more elasticity around the low points in business activity when uncertainty and bond market speculation are rampant. Such a point was reached in 1958, and, for this reason, analysis of the 1957-58 recession may yield an explanation for the pattern of residuals.

9. The 1957-1958 Recession[1]

The 1957-58 recession differed from its predecessor in several important respects. Interest rates in the summer of 1957 were the highest in many years; Standard and Poor's Index of Stock Prices had more than tripled since 1946 with more than 70% of the rise having occurred since January 1954; the Consumer Price Index, except for pausing in 1953, 1954 and 1955, had risen ten points during the decade with the Wholesale Commodity Index following suit; the average maturity of the Government Debt had shrunk from five years and six months in 1954 to four years and nine months in 1957 and, much more important, the Treasury Department was determined to lengthen it; there had been a fortuitous tax cut in 1954; the financial community had expected the Federal Reserve to follow an easy money policy in 1954 and the monetary authorities had obliged—in 1958 the financial community expected the same policy only to be disappointed by an initial hesitation on the part of the "Fed" to take positive steps toward monetary ease followed by a cheerful readiness to take positive steps toward monetary restraint at the first sign of recovery; and finally, the 1957-58 drop in economic activity was much sharper and of shorter duration than in 1953-54. Consequently, the pressures on the bond market were too strong and changing too swiftly to be accommodated without occasional, temporary deviations from the "normal" bond yield differentials.

The bond market was in a highly volatile condition entering 1957. "Events during the previous recession had provided ample evidence that market rates of interest could be expected to move downward with general economic conditions, as a result both of changes in private demands for funds and of shifts in monetary policy. Investor expectations were so sensitively attuned to these relationships that on earlier occasions during the cyclical advance of 1954-57—most notably in early 1957—interest rates had declined and bond prices had experienced a sharp temporary upward movement set off by transient indications of a slowdown or possible reversal in the pace of business activity."[2] Normally inelastic expectations were experiencing

1. For the ensuing analysis information concerning the general condition of the bond market has been gathered from: The Commercial and Financial Chronicle, Barron's, Treasury-Federal Reserve Study of the Government Securities Market (Washington, 1959), and Joint Economic Committee, A Study of the Dealer Market for Federal Government Securities (Washington: U.S. Government Printing Office, 1960).

2. Treasury-Federal Reserve Study, Part II, p. 3.

some strain awaiting a signal which would indicate a turn in the bond market. This condition is evidenced in the pattern of residuals in January and February 1957 before the dip and, again, after the recovery in November and December 1958. In both instances there are indications of a plus, minus sequence as false starts result in overstated risk premiums subject to overcorrection as investors realize their mistake and confidence returns.

In the summer of 1957 there were scattered indications that the observed differential was tending to exceed the computed differential, i.e., there were signs that pessimism was building even though general business indicators remained high. Another factor was that the speculative excess of the long capital spending boom had not been corrected and a general desire to invest in competing long term assets prevailed. There were a few signs that the economy was weakening, but the demand for loanable funds was very strong and interest rates were high. Some banks and institutions were switching from long term Governments and Aaa Bonds to short term Governments due to rising uncertainty, while others were selling, especially longer term Aaa's to obtain funds for loans.[1] Added to this was the selling pressure from investors who were moving from bonds to stocks.[2] The combined effect was a higher differential than "normal" between Corporates and Governments and a lower differential than "normal" between Aaa's and Baa's as selling pressures from all sides descended upon higher quality Corporate Bonds. By June the spread between Corporates and Governments was the highest in many years, and uncertainty concerning the economy was building up. In July some buying appeared in bonds, but it was mostly of a professional nature. In August the stock market turned down and unemployment began a slow rise. At the same time the prime rate for bank loans and the rediscount rate were raised leading to a great deal of uncertainty and confusion. More money became available for bonds and the Government market stabilized; but the desire for liquidity was strong and selling of Corporates continued although trading volume fell off.

In September and October the situation changed. The computed differential basically exceeded the observed differential in these two months, indicating a high degree of confusion in the money market and supporting the

---

1. For example, in the second quarter of 1957 Commercial Banks sold 4.5 billion dollars of long term (over one year) Federal Obligations and increased short term holdings by 2.2 billion and loans by 2.1 billion. Source: Federal Reserve Board, Flow of Funds/Saving Estimates, Supplement No. 4 (Washington, 1961).

2. This could be stated alternatively as the premium necessary to prevent bond holders from indulging in portfolio shifts. Consideration of this distinction would involve detailed microanalysis utilizing data not readily available at present. Moreover, it will later become clear that portfolio shifts among existing assets are far less important in the long term (over 10 years) bond market than the absorption rate for new Government Bonds.

contention that interest-expectations tend to be inelastic until "significant reason" is given for a change. Business conditions were worsening, but without some specific confirmation of their fears investors were not evaluating risk accordingly. There had been false starts before (as in January), and investors weighed the risk of capital loss, should a sudden business upturn develop causing bond prices to fall, against the probability of capital gain should a true business downturn occur. There was some buying. Smaller commercial banks, experiencing a drop in the demand for loans, began buying Aaa Bonds for income purposes, and sophisticated investors started moving from stocks to bonds, particularly Baa's. The Government market remained relatively weak although a 4% twelve year bond issued for cash was readily absorbed during September.[1] During these two months, the professional who, in this thin market, carries out arbitrage transactions in accordance with his estimate of the proper interest rate structure was giving away to the investor throughout the market.

On November 14 the "significant reason" was given, and the elasticity of expectations with respect to the interest rate became immediately negative as a downturn in business was confirmed. On that date the rediscount rate was lowered by the Federal Reserve System. "By far the most important single influence on interest rate expectations, therefore, is the outlook for general business conditions. Signs of changes in monetary policy are watched closely by participants in the Government securities market since such changes are often interpreted as confirmations of major turns in the business situation and suggest to observers possible shifts in the balance of pressures in the money and capital markets."[2] The observed differential greatly exceeded the computed differential as the risk premium adjustment started with a rush. The basic yield adjusted first, falling markedly, with Aaa's close behind. Initially the lower quality yields rose; longer term Baa's were sold heavily as risk evaluation among monetary assets outweighed any evaluations based upon the choice between equities and fixed income bearing securities. Although the imperfect nature of the bond market precludes immediate adjustment, its tendency to maintain a "normal" differential soon reasserted itself, and, by the beginning of 1958, most risk premiums reflected existing conditions.

During the Spring of 1958, as bond prices rose, expectations tended to have an elasticity greater than one, i.e., every change in interest rates was interpreted as indicating that the trend would continue. In January a second

1. One would expect this new issue of Government Bonds to tend towards a narrowing of the differential. However, the lack of correlation between particular new issues and the dependent variable is indicated by the signs in the pattern of residuals in Table 2 at the time of this issue and of three other cash issues: minus, plus, plus, "normal" for 9/57, 11/57, 2/58, and 6/58 respectively.

2. Treasury-Federal Reserve Study, Part II, p. 3.

cut in the rediscount rate and a cut in margin requirements were reinforced by the more positive fact of a cut in the prime lending rate by commercial banks. Through the Spring the rediscount rate continued to fall, reserve requirements were cut and, again, the trend was confirmed by an April cut in the commercial banks' prime loan rate. All this caused investors to seek "certain" capital gains in the bond market as rising bond prices were extrapolated well into the future. A large speculative element arose aided and abetted by the determination of the Federal Government to lengthen the maturity structure of the national debt. As bond prices rose, the risk of capital loss due to a business upturn increased. The premium demanded for holding lower quality bonds showed a tendency to exceed that warranted by existing conditions. This tendency of the observed differential to exceed the computed differential for Baa Bonds became more pronounced as the stock market launched its recovery in March. Selling pressure from this source alone might have been absorbed within "normal" limits, but speculative buying was heavy in the high quality market encouraged in some instances by brokerage houses. Expectations were for an extended period of monetary ease and recommendations were being made for the purchase of long term Government Bonds for the recession and also of Aaa's if the yield spread were 50 basis points (.5%) or more. Hypothecating at the bank was presented as a method of maintaining actual cash investment relative to the total value of holdings at a minimum thus enhancing profit potential.

Taking advantage of excess demand, suppliers provided new bond issues at a rapid pace. Since it is the Government yield curve which serves as the base from which risk premiums are calculated and indications are that differentials, except under stress, are predominantly a function of existing economic conditions, review of changes in Government Bonds outstanding during the 1957-58 recession should provide some insight into factors affecting the over-all interest rate structure.[1] In February the Treasury sold six year and thirty-two year bonds in an exchange operation. This financing raised the amount of bonds (as opposed to bills, certificates and notes) represented in this fifteen billion dollar segment of Government debt from 1.5 to 4.5 billion dollars. There followed, in the same month, the sale of a nine year bond for cash. These issues, coupled with the twelve year bond of September and the seventeen year bond of November 1957, both for cash, increased the value of Government Bonds outstanding with over ten years to maturity by approximately 3 billion dollars. This set the stage for the speculative blow off which took place in conjunction with the final attempt of the Treasury to lengthen the debt with its June financing consisting of a 1.1 billion dollar twenty-seven year bond for cash and a 7.4 billion six year eight month bond in exchange for existing securities of predominantly shorter maturities. Of the 4.1 billion addition to Government Bonds of over

1. All data concerning Government Bonds has been taken from the Treasury Bulletin.

ten years maturity between June 1957 and June 1958, 25% was absorbed by commercial banks and 50% by investors other than banking and financial institutions. In fact, absorption of new issues accounted for the only portfolio changes that occurred with respect to Government Bonds with maturities greater than ten years:

(billions)	Total	Fed. Res.	Comm. Bks.	Sav. Bks.	Insur.	Other
June 1957	30.8	1.4	4.2	4.0	4.6	13.7
June 1958	34.9	1.4	5.2	4.1	5.1	15.9
Net Change	4.1	(+.03)	1.0	0.1	0.5	2.2

It has been stated that new issues have no statistical significance when included as an independent variable in the regression equation. The facts that the timing of new issues is carefully planned and, in the case of long term Treasury Bonds, restricted to periods when demand is high make this result credible. However, there is an impact to be reckoned with. New issues were absorbed by the more volatile elements in the market—commercial banks and "other" investors. Commercial banks expected to hold long term bonds only until loan demand revived and were anxious to take profits six months after purchase. "Other" investors were seeking an alternative source of capital gains—a "sure thing" with falling interest rates—until such time as an upturn in business was confirmed bringing with it a return to tight money policies. Up to May expectations were almost universally held (with "certainty") that interest rates would continue to fall as the Monetary Authorities attempted to stimulate a slowly recovering economy. In late Spring 1958 a sort of Damocles' sword overhung the market waiting for expectations to change before dropping. "In the area of debt management, there is the problem as to whether, in periods when easy credit conditions lend investor favor to longer term, higher yielding issues, a large and rapid shift in the maturity structure of the debt may result in supply and demand distortions, which may later have upsetting and disruptive effects on the market." [1]

The Treasury's June financing attracted a large amount of speculative funds. The failure of this issue to provide immediate capital gains was strong evidence that expectations were outrunning the actual course of events. As distress selling occurred, the Treasury and the Federal Reserve found it necessary to support the long term market in June and July. Signs were growing that business recovery was underway. By May the stock market was well on the road to recovery, and the selling of Baa's is clearly reflected in the large differential between Baa's and Governments and between Baa's and Aaa's. Bank loans were picking up and selling pressure was building in the Corporate market. The Government market was in a technically weak position with speculators desiring to sell and institutional portfolios filled. In

1. Treasury-Federal Reserve Study, Part II, p. 102.

July it became evident that future Treasury financing would be short term, and this was taken as a clue that further cuts in the rediscount rate would not occur. Now, expectations reverted to their normal, inelastic state. Although it was fairly clear that interest rates would not fall further, it was unclear to many that they would reverse direction in the face of the continuing high level of unemployment and the anticipated slowness of the business recovery. Once again confirmation of a change was required before expectations exhibited negative elasticity.

In August the computed differential generally exceeded the observed differential as a congeries of factors added up to a "significant reason" for a change in expectations. The short term interest rate on prime commercial paper turned up indicating a reduction in liquidity; the stock market was rising. Margin requirements were raised, and it became clear that, in Washington, inflationary fears were taking precedence over concern with the economy. Those who had participated in the February Treasury financing had passed the six month barrier. The sword fell. To absorb the increased supply of Government Bonds a severe adjustment in the basic yield was required. The bond market initially was unable to accommodate the rush of selling pressure.[1] Although professionals believed the market was overdiscounting any future change in credit conditions, institutions and dealers were unwilling to commit funds until the market had stabilized. Banks which had bought in June suddenly found themselves with short term capital losses which they were unwilling to realize until they became long term at the end of the year or early in 1959. Within the next two months the market did adjust, and the yield differentials reverted to "normal" limits. Just as November 1957 marked the beginning of the Recession Phase in the bond market as expectations of the "average" investor with respect to bonds turned bullish and everyone became a "buyer," so did August 1958 mark the end of this phase as expectations turned bearish and everyone became a "seller." "Basically both these periods of accelerated price adjustment were reflections of the high sensitivity of the Government securities market in recent years to shifts in the direction of economic activity and of the strategic role played in this pivotal market by the expectations of participants. In the summer of 1958, the timing of the economic upturn as well as the speed with which recovery in activity took hold was a surprise to the market. As evidence continued to pile up that recovery was

---

1. It is interesting again to note that selling pressure can be stated equally well as that rise in interest necessary to make investors content with their existing portfolios. Ownership figures for Government Bonds indicate no portfolio shifts among existing securities during this entire adjustment period. Conard, p. 315, indicates this to be a somewhat normal condition as investors throughout the market adjust prices with respect to changes in expectations arising from any segment of the market without a great volume of shifting.

proceeding, the impact on expectations was pronounced and the adjustment of Government securities prices severe." [1]

10. Conclusions

There are three basic approaches to the explanation of the interest rate structure involving:

(1) Institutions
(2) Loanable Funds
(3) Expectations

In this study of bond yield differentials the last approach has been favored with emphasis upon the subjective evaluation of risk under conditions of uncertainty. However, throughout the discussion frequent reference has been made to factors which may readily be classified under the general headings of either loanable funds or institutions. To justify the emphasis upon expectations and to provide clear cut conclusions, it is necessary to review the three possible approaches in the context of the results achieved. Since we are interested in the dynamics of interest rate structure, a period of maximum change in prices, yields and differentials would seem to be best for illustrative purposes. Such a period was July 1957 through February 1958—the month preceding the stock market peak to the month preceding the stock market recovery.

Table 5 shows the change in basis points attributed to each independent variable by the regressions and the net estimated change of the differential over the period. It is evident from the table that:

(a) differentials widened during this recessionary period
(b) lower quality differentials widened more
(c) changes in Corporate volume outstanding offset to a large extent, but not entirely, the impact of changes in Governments outstanding and both showed a slightly greater impact for longer maturities
(d) of the three remaining variables: changes in equity yields have by far the greatest impact; changes in short term interest rates have less impact as maturity lengthens; and economic conditions are much more important in the B than in the A equations.

1. Treasury-Federal Reserve Study, Part II, p. 78. The increasingly high sensitivity to the direction of economic activity may readily be explained by the secular shift in the composition of ownership of U.S. Securities. For example, between 1954 and 1959 total long term Government debt outstanding fell 22% with percentage holdings (other than Federal Reserve System, Government Agencies and Trust Funds) as follows:

	Commercial Banks	Savings Banks	Ins. Co's.	Other
January 1954	15%	16%	20%	37%
December 1959	10%	8%	14%	54%

The more volatile elements—commercial banks and "other" investors—thus have become a much more important factor in the market. Particularly in 1958 this led to instability. Source: Treasury Bulletin.

TABLE 5[1]

Differential Changes Attributed to Specific Independent Variables
July 1957 to February 1958

	Estimated Net Change in Differential Due To:		Corporates Long Term Debt	Governments Long Term Debt	Interest Rate	Equity Yield	Unemployment
Change in Independent Variables			+ 3.0	+ 2.3	- 1.25	+ .61	+ 2.5
Basis Point Change (Y_c)				(net) (1 & 2)			
A-12	+21.1	=	+28.4	- 31.5 (- 3.1)	+ 8.4	+15.1	+ 0.7
A-16	+24.0	=	+40.0	- 50.0 (-10.0)	+ 7.0	+17.0	+10.0
A-21	+17.0	=	+51.0	- 58.0 (- 7.0)	+ 3.0	+15.6	+ 5.4
B-12	+93.0	=	+68.0	- 98.5 (-30.5)	+28.0	+64.0	+31.5
B-16	+75.0	=	+95.0	-120.0 (-25.0)	+20.0	+50.0	+30.0
B-21	+63.0	=	+96.0	-118.0 (-22.0)	+ 8.0	+46.0	+31.0

1. See notes on following page.

Notes for Table 5:

Actual differential changes, i.e., changes in the dependent variables were:

	A	B
12	+ 35	+ 110
16	+ 16	+ 79
21	+ 4	+ 61

In the cases of A-12 and A-21 the divergences between ΔY_C and ΔY_O are marked. In the other cases the same general magnitude prevails. This is the result of choosing a single trading day in a particular month during a period in which an imperfect market is being subjected to the stress of highly volatile expectations. All the adjustments in evidence at any point in time will not be one hundred per cent successful. For example, A-12 through January 1958 showed $\Delta Y_C = +21$ and $\Delta Y_O = +19$; A-21 through March 1958 showed $\Delta Y_C = +20$ and $\Delta Y_O = +15$—discrepancies well within normal limits. On the other hand, choosing March 1958 as the terminal date would have shown for: B-12, $\Delta Y_C = +95$ and $\Delta Y_O = +129$; B-21, $\Delta Y_C = +67$ and $\Delta Y_O = +46$—discrepancies which appear rather large. In any case the general conclusions utilizing Table 5 do not appear to be invalidated by an argument to the effect that observed yields acted in any consistent manner differently than computed yields. As the preceding analysis has demonstrated, the differentials do maintain a "normal" relationship over time subject to occasional, temporary displacement. Therefore, use of the "normal" differential is justified.

TABLE 6[1]

Patterns of Prices, Yields and Differentials
July 1957 to June 1958

A. 1970 Maturity

Coupons: 3-3/8 Baa (adj.)
3 Aaa
2-1/2 Government

1957	Baa Yield	Baa Price	Aaa Yield	Aaa Price	Government Yield	Government Price	Diff.-B	Diff.-A
							(Basis Points)	
July	4.47	89-1/4	3.99	90	3.82	86-1/2	65	17
Aug	4.55	88-1/2	4.10	89	3.76	87	79	34
Sept	4.95	85	4.18	88-1/4	3.85	86-1/4	110	33
Oct	5.06	84-1/8	3.99	90	3.83	86-1/2	123	16
Nov	5.10	83-3/4	3.94	90-1/2	3.24	92-1/4	186	70
Dec	5.49	80-1/2	3.52	94-1/2	3.07	94	242	45
Jan	4.92	86	3.28	97-1/8	3.08	94-1/4	184	20
Feb	4.96	85-3/4	3.33	96-3/4	2.95	95-1/2	201	38
Mar	5.09	84-3/4	3.32	96-3/4	2.85	96-1/2	224	47
Apr	5.16	84-1/8	3.23	97-3/4	2.87	96-1/4	229	36
May	5.17	84	3.14	98-5/8	2.86	96-1/4	221	28
June	5.01	85-3/8	3.31	97	3.05	94-1/2	196	26

1. See notes on p. 44.

TABLE 6 (continued)

B. Means Over Ten Years

Average Coupon and Maturity:

4	1978	Baa (adj.)
3	1979	Aaa
2-5/8	1971	Government

1957	Baa Yield	Baa Price	Aaa Yield	Aaa Price	Government Yield	Government Price	Diff.-B	Diff.-A
July	4.90	88-1/4	4.12	84	3.66	88-7/8	124	56
Aug	4.99	87	4.18	83	3.68	88-3/4	131	50
Sept	5.09	86	4.16	83-3/8	3.65	89	144	51
Oct	5.20	84-3/4	4.12	84	3.70	88-5/8	150	42
Nov	5.27	83-1/2	4.02	85-1/4	3.41	91-3/8	186	61
Dec	5.13	85-1/2	3.68	89-7/8	3.24	93-1/8	189	44
Jan	4.92	88-3/8	3.60	91-1/4	3.09	95	183	51
Feb	4.93	88-1/4	3.61	91-1/8	2.99	96-1/8	194	62
Mar	4.99	87-1/2	3.65	90-1/2	2.90	97	209	75
Apr	4.97	87-3/4	3.51	92-1/2	2.89	97-1/8	208	62
May	4.96	87-7/8	3.53	92-1/8	2.88	97-1/4	208	65
June	4.86	89-1/8	3.59	91-3/8	3.06	95-3/8	180	53

Notes for Table 6:

(1) All prices in this table have been approximated according to a standard bond value table.
(2) In Section A the differentials shown are not those used as the dependent variable in the regression analysis. The dependent variables were calculated from pairs of curves fitted to yield scatters.* The above calculations are based upon actual yields utilizing the average of two Baa Bonds and single Aaa and Government Bonds.
(3) In Section B yields are from the regression data for means over ten years.
(4) In the case of Baa Bonds the yields are after adjustment as described in the Appendix. (It is interesting to note that even without the adjustment the Baa data exhibit the same general price and yield pattern.)

* Actual Differentials Used as Dependent Variables for:

	B-12	A-12
July	86	20
Aug	106	24
Sept	117	17
Oct	129	15
Nov	195	67
Dec	201	38
Jan	175	39
Feb	196	55
Mar	215	55
Apr	208	33
May	211	35
June	181	28

Table 6 shows monthly prices, yields to maturity and differentials involving the three bond categories for means over ten years and the maturity year 1970. Only one maturity year is shown since the regression results all yield the same general trend with only variations in magnitude and since, for this maturity, there existed actual bonds outstanding in each category; checking the 1970 pattern against that for other points along the yield curves a high degree of similarity is found. In the case of the means (which do not have exactly the same average maturities) the adjustments in prices and yields are somewhat smoother, the recovery of B bonds takes place one month earlier, and it is more complete. Since we are interested merely in illustrating the general trend, the data shown should be adequate. It is evident from this table that:

(a) A Bond and Government Bond prices rose considerably between July 1957 and February 1958
(b) B Bond prices were not greatly changed by February 1958 due to an initial fall followed by a mild recovery toward the end of 1957
(c) the greater change in the lower quality differential was associated with the relative stability of B Bond prices.

The implications of this experience with respect to debt management have been fairly well explored in the earlier analysis. Tables 5 and 6 confirm the previous assertion of the importance of Treasury financing and its potentially significant impact. The fact that all issuers of fixed interest bearing securities tend to follow the same borrowing pattern creates an offsetting situation in so far as the differential is concerned, albeit only as a result of the awareness of the protagonists. With respect to monetary policy Tables 5 and 6 do provide some additional insight. Table 5 shows that the commercial paper rate, the most sensitive to Federal Reserve action, dropped 125 basis points between July 1957 and February 1958. Section B of Table 6 shows that the mean yields for Governments, Aaa's and Baa's changed by −67, −51 and +3 basis points respectively. In other words, while high quality bond prices rose approximately $7, lower quality prices remained practically unchanged (following an initial drop). This fact would tend to support the argument that an easy money policy benefits primarily the larger corporations which enjoy the highest credit ratings as opposed to the more nearly marginal firms for which the need under adverse circumstances may actually be greater. The Aaa corporation can go to the market at a cost reduced from 4.12% to 3.61%; the Baa corporation, on the other hand, must still pay in the neighborhood of 5% to attract funds on the long-term bond market. It seems evident that quality, i.e., risk factors, outweigh factors bearing upon the structural inter-relationship of yield curves approached in a purely mechanistic manner. The above basis point changes indicate a relationship between the short and long term rates; the high quality short rate moved lower and with it the long but the latter to a much smaller extent as adverse risk evaluation weighed more heavily. Therefore, open market operations appear weak when judged upon the basis of their impact on long term rates. The data also indicate that they are weak upon an inter-quality

basis, given the same maturities. The former problem can be met by extending the maturity range of Federal Reserve purchases, but, if expectations are adverse, the latter problem involving favored treatment for the best credit ratings remains—except in so far as Baa concerns might be able to benefit from a generally more effective lowering of long term borrowing costs. It seems that changes in yield differentials can affect the potency of monetary policy even barring any problems involving maturity. It is, therefore, important to understand what is behind them. If it is expectations, it seems dubious that the purchase and sale of Government Securities can, alone, provide an interest rate structure acceptable to all.

Can these changes be explained by institutional factors?

Certainly the actions and statements of the Federal Reserve Board are among the most important influences in the bond market. Its actions can influence interest rates and the availability of credit thus having a decided impact upon bond prices, but there is no explanation here of how this impact is distributed among various bond categories especially in the long term market. As long as the Federal Reserve adhered to "bills only," it does not seem feasible to assign to it direct responsibility for changes in the interest rate structure for maturities greater than ten years. Only in June and July 1958 when direct support of the long term market was undertaken can direct responsiblity be assigned. Otherwise, the influence of the monetary authorities is felt indirectly through its effect on expectations as to the future course of interest rates.

Another potentially important institutional argument hinges upon the fact that many financial institutions are restricted to Government, Aaa and some Aa Bonds. But if this explanation is used to explain the greater B bond differential, what is left to explain the fact that the A bond differential widens and narrows in accordance with changing conditions in the economy and in the financial market? For many investors there is no such restriction upon the quality of bond they may purchase. If these investors believed the yield differential so wide as to compensate for the additional estimated risk, there is no reason why arbitrage transactions would not take place.

Quality ratings in themselves constitute an institutional fact, but they merely indicate that a differential should exist not how it may vary under changing conditions. The question to be considered is why the differential between two quality ratings is changed at various points in time, and this question comes down to the consideration of risk and the compensation demanded for undertaking it. Institutional factors may temporarily influence the results at any point in time but can hardly explain them.

Can these changes be explained by supply of and demand for loanable funds?

The equation used in this study can be interpreted in the following manner. The first two independent variables, the ratios of Corporate and Government Bonds to Long Term Debt, represent the supply of bonds (relative to competing assets). The last three variables can be interpreted in terms of the supply of and the demand for money. The short term interest

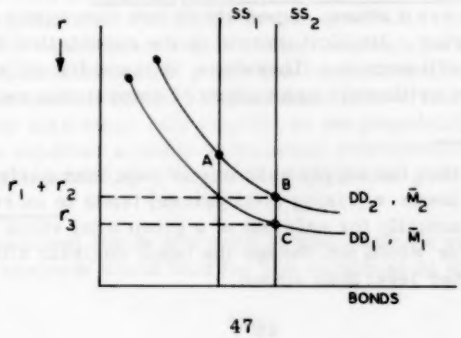
rate (r) serves as an indicator of the supply of loanable funds or liquidity in the money market. The yield on equities (E) serves as an indicator of the demand for loanable funds or the demand for real assets, i.e., a proxy for investment demand. The rate of unemployment (U) serves over a short interval as an indicator of transactions demand due to changes in the level of national income. If we think of a three asset world having only money, bonds and capital, we can represent a change in the amount of money (M) available either to be held as cash balances or to be exchanged for bonds as the algebraic sum of changes in the three uses and sources of funds outlined above. Let us designate them as follows:

(a) M_M = supply of loanable funds = f(r)
(b) M_E = net demand for loanable funds = f(E)
(c) M_T = demand for transactions funds = f(U)

A change in any of the parameters will either result in an increase or decrease in the amount of funds (M) available for holding as monetary assets, i.e., cash or bonds. The designation of M_E as "net" means that any change in demand for real assets resulting from changes in M_M or M_T is automatically adjusted for, leaving:

$$\Delta M_M - \Delta M_E - \Delta M_T = \Delta M$$

Between July 1957 and February 1958 r fell indicating an increase in the supply of loanable funds, E rose indicating a decrease in the demand for loanable funds and U rose indicating a decrease in transactions demand. If we assume that a portion of the increased M funds will be used to purchase bonds, since all the above changes add to the supply of M, the demand for bonds must have risen. At the same time the volume of bonds outstanding rose, i.e., supply rose. There is no denying the importance of the supply side in analyzing the bond market. However, it is almost a passive element or a function of the demand for bonds; every effort is made to change the supply in such a manner as just to absorb the available funds without upsetting the existing rate structure. If the borrowers were absolutely successful, there would be no change in the interest rate at the time of increased supply. If demand falls back to its original level thereafter, an adjustment in interest rates (upward) will take place as illustrated in the diagram below:



The horizontal axis measures the volume of bonds in the conventional manner; the vertical axis, however, represents the rate of interest rising as it approaches the origin. A is the initial equilibrium; B is the equilibrium attained by an exact matching of supply to the increased demand; and C is the equilibrium following a return of the demand schedule to its original position. [1] The final equilibrium is at a higher rate of interest (lower price) than originally and illustrates in greatly simplified form the type of adjustment in the bond market which took place during 1957–1958.

What has the above analysis told us? It has told us that under circumstances similar to those between July 1957 and February 1958 people would increase their demand for monetary assets. But it has told us nothing of the choice among closely substitutable assets—between money and bonds, between different categories of bonds. We have slipped back to the interest rate and, more importantly, have had to assume that an increased demand for monetary assets is equivalent to an increased demand for bonds. This latter assumption is not necessarily true nor can it be expressed as some mechanical function of the funds available for monetary assets (although the assumption is usual enough to serve as a first approximation). To illustrate, assume an increase in funds available for aquisition of monetary assets under the following conditions:

(1) a. Investors expect equity prices to fall
    b. " " bond " " rise
(2) a. Investors expect equity prices to fall
    b. " " bond " " fall.

In the first case the tendency would be to buy bonds and sell equities, but this may be tempered, or upset, by a tendency to shift into cash in the event overall expectations concerning general economic conditions are sufficiently pessimistic. In the second case, e.g., a panic following speculative excess, the tendency would be to buy neither bonds nor equities but to shift into cash. Thus, one cannot say whether or, more especially, how bond prices will be affected by an increase in funds available for monetary assets. In order to explain interest rate structure, one must be concerned with the allocation of funds among closely substitutable assets; and in order to explain this allocation, one must bring in expectations and the subjective evaluation of return and risk under conditions of uncertainty.

Can these changes be explained by expectations?

Returning to case 1 above, expectations are that equity prices will fall and bond prices will rise. Implicit therein is the expectation that general economic conditions will worsen. Therefore, demand for different bond categories is not a simple arithmetic calculation of comparable returns for various

1. It may be argued that the supply schedule is less than perfectly inelastic due to the fact that a lower r (rising bond prices) tends to increase the "floating" supply of bonds actually for sale out of a given total stock oustanding. Allowing for this factor would not change the basic analysis although r_3 would be at a somewhat higher level than shown.

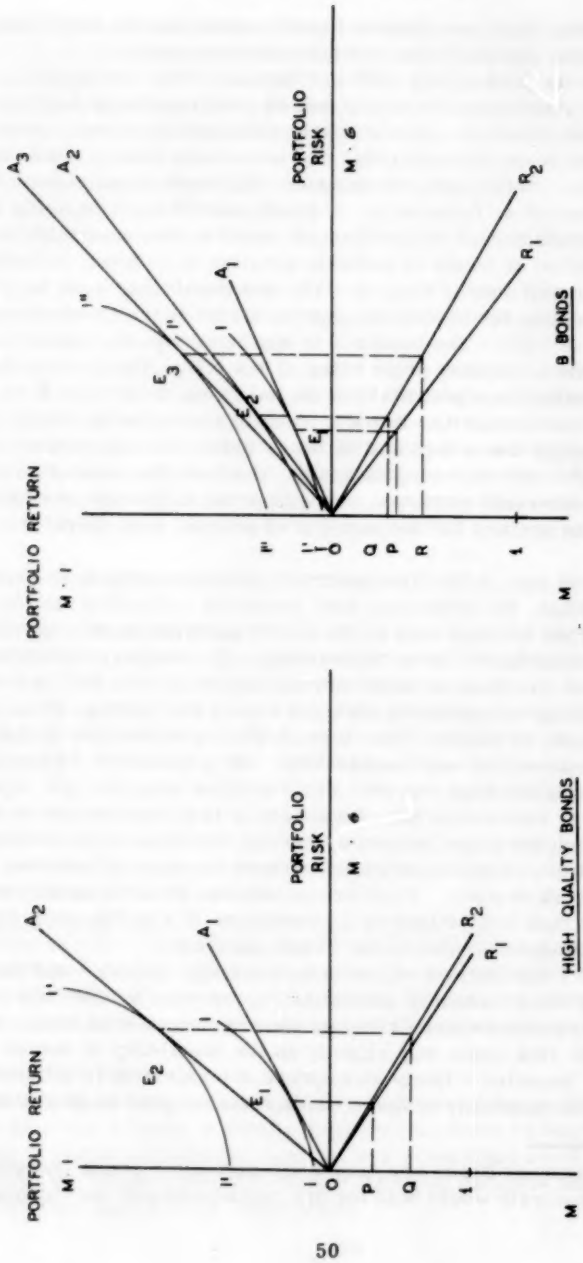
qualities but depends also upon personal preferences and the subjective evaluation of risk. Bonds are preferred but not indiscriminately.

Looking at the period July 1957 to February 1958, probabilities favored a relatively short run recession as opposed to a depression of long duration. These expectations, however, were not held with certainty, and, therefore, there was an increase in estimated risk and a tendency toward bonds offering safety and liquidity. After cash, Government (G) Bonds come closest to having these characteristics, followed by A bonds with B bonds running a sorry last. Thus, we would expect the yield on B bonds to increase relative to A bonds and the yield on A bonds to increase relative to G bonds as reflected in the price and yield patterns of Table 6. The end result may even be a cumulative process accounting for speculation, particularly in the Government sector, such as occurred in 1958. The tendency to buy higher quality bonds in a period of uncertainty leads to relative price rises $G > A > B$. The pattern of price rises, in itself, indicates a potential for capital gains in $G > A > B$ and, therefore, attracts speculation in $G > A > B$. Table 6 shows the divergent price pattern for higher quality bonds (up) and B bonds (down then up) between July 1957 and February 1958. The following diagrams illustrate the manner in which, given individual preference patterns, changes in the subjective evaluation of risk and return can account for the patterns of prices, and, therefore, of yield differentials.

The vertical axis in the first quadrant measures estimated return for various bond holdings, the horizontal axis measures estimated risk for various bond holdings and the vertical axis in the fourth quadrant measures bond holdings. OA is the opportunity locus representing all possible combinations of risk and return for the stock of bonds outstanding from zero risk and return for zero bond holdings to maximum risk and return for holdings equal to the total available supply of bonds. The slope is i/σ, or return per unit of risk, and is assumed constant for any combination. OR represents all possible combinations of bond holdings and risk with the slope equal to $1/\sigma$, again assumed constant. II represents one of a family of indifference curves between risk and return with the slope indicating that any increase in the risk content of portfolios must be accompanied by an increase in expected returns, i.e., the investor is a risk avoider. Equilibrium between investor preferences and return per unit of risk is indicated at a tangency of II and OA, with the bond content of the portfolio indicated in the fourth quadrant.

In July 1957 equilibrium exists in both the high quality [1] and Baa categories at E_1 with bond content of portfolios represented by OP. As the recession begins, estimated return rises as bond prices start to rise. In the high quality sector risk rises only slightly as the possibility of default is very small even if one expected a fairly widespread deterioration in economic conditions, and the marketability of these bonds remains good as investors attempt

1. Aaa and Government Bonds are here combined into a single category although the same analysis would hold for any combination of two categories.


PORTFOLIO RISK
M · σ
PORTFOLIO RETURN
M · I
A3
A2
A1
I'''
I''
I'
I
E3
E2
E1
R2
R1
I''' I'' I' O Q P R t
M
B BONDS
PORTFOLIO COMPOSITION
PORTFOLIO RISK
M · σ
PORTFOLIO RETURN
M · I
A2
A1
I'
I
E2
E1
R2
R1
I'' I O P Q t
M
HIGH QUALITY BONDS
PORTFOLIO COMPOSITION
M = BOND HOLDINGS
I = ESTIMATED YIELD
σ = ESTIMATED RISK

to acquire a return at minimum risk. With the increase in $i > \sigma$ the slope of OA (i/σ) rises and a new equilibrium is reached at E_2. The slope of OR $(1/\sigma)$, on the other hand, does not change greatly. The proportion of bonds desired in the portfolio therefore rises to OQ with the result that the high quality bond prices are bid up. As indicated above, the increase in expected return may become cumulative, and as long as it does, prices should tend to rise.

In the B bond sector results are somewhat different. Estimated return has risen but so has estimated risk. The possibility of default is more real as indicated by the lower quality rating. Ratings are based upon such factors as greater fluctuations in earnings and less ample coverage of fixed charges by those earnings. This combination couples a lower tolerance for instability with a higher degree of it meaning that even taking safety alone into consideration the estimated risk would rise more for B than for higher quality bonds. But in addition to less safety there is less liquidity. Between July 1957 and February 1958 investors were putting money into the bond market as a temporary measure. To them, liquidity may be defined as the expectation of a rising or at least a stable price during the investment period. The estimated risk on this score is also greater for B bonds. Thus, the slope of OA (i/σ) may increase if i increases more than σ but not by as much as in the high quality category. The slope of OR $(1/\sigma)$, on the other hand, falls more for the B bonds with the result that the new equilibrium at E_2 indicates a decrease in the desired proportion of B bonds held from OP to OQ. The tendency then is for B bond prices to fall. If estimated return thereafter rises sufficiently, equilibrium at E_3 may be reached with the portfolio content of B bonds rising to OR resulting in increasing bond prices. This type of demand analysis affords the only adequate explanation of the manner in which wider differentials were determined between July 1957 and February 1958.

We have now completed both a quantitative and qualitative analysis of bond yield differentials. The present and the preceding sections have supported the contention that an expectational approach to the theory of the interest rate structure is required. Although this qualitative analysis has been based almost exclusively upon the 1957–58 recession, the relative emphasis should not be thought to detract from the generality of the conclusions. The period was chosen for expository purposes with the knowledge that the forces to be explained are in particular evidence during sharp changes in the level of economic activity and, therefore, more readily pinpointed. That such changes, whether more or less sharp, occur with sufficient frequency to allow generalization of these qualitative conclusions is self evident. Turning to the quantitative conclusions, a brief review may be in order.

Contrary to the original hypothesis it was found that independent variables representing the following were not significant:

(1) the relative volume outstanding at specific maturities
(2) changes in the relative volume outstanding at specific maturities and
(3) new bond issues.

Within the general category of stock and flow variables the only significant

results were obtained from the relative volumes of Government and Corporate Bonds, respectively, and competing long term debt. The broad conclusion is that the traditional representation of yield curves in the long term bond market as smooth, continuous functions is a close approximation of the real world. No support was evident for the hypothesis that the actual function is discontinuous and/or subject to humps depending upon conditions at specific maturities. The above must be qualified, as mentioned often in the previous discussions, by the fact that new issues are handled gingerly by both the private and public sectors. A sudden change in this condition might have material impact upon the yield curves—with present knowledge, of an imponderable magnitude.

Of the expectational variables the equity variable showed by far the strongest impact. Given the relative volume of bonds outstanding, conditions in the money market, capital market and the economy in general offer a basis for the explanation of a large percentage of the variation in bond yield differentials. The relationship is a consistent one not significantly altered by cyclical factors per se, although it cannot be gainsaid that cyclical elements are incorporated in the fluctuations of each of the expectational variables. More specifically, the money market variable (the rate on prime commercial paper) showed declining strength and an insignificant relationship as maturity lengthened. This result tends to support the traditional view that the impact of the short term interest rate upon the over-all structure weakens with maturity thus contributing to a greater stability in the very long term relative to the shorter terms. It is also interesting to note that in the high quality, highly sophisticated markets changes in the interest rate structure anticipate changes in the short term rate. The variable representing general economic conditions and at the same time serving as a proxy for the subjective measure of default risk (the unemployment rate) showed an insignificant relationship for the A equations and a significant, consistent one for the B's. Since Government and Aaa quality ratings are the same, such a result is easily credible. The consistency of the coefficient values for the B equations indicates that default risk looked upon as a function of general economic conditions (and therefore of corporate earnings) is not affected by maturity, i.e., the risk of default is no more nor less for an intermediate than for a long term bond. If a corporation defaults, all its bonds are equally affected. As mentioned above, the strongest expectational variable, and the one which exhibited significance throughout, was the capital market variable (equity yield). For the A equations, involving bonds of the same quality rating, there was no definite trend in the coefficient values. For the B equations a definite decrease in coefficient values was evident for longer maturities. Thus, there seem to be two elements in the decision as to the risk content of existing portfolios. When the investor seeks safety, for example, he tends to switch from physical, or capital, assets to monetary assets, but at the same time he tends to make a choice within the range of monetary assets with respect to quality and maturity. On the average investors seeking safety attach somewhat higher priority to liquidity and the absence of default risk, i.e., there is a preference for higher quality, shorter term bonds as well as for bonds as such. The result is

reflected in the pattern of coefficient values as maximum impact occurs at the short end of the maturity range.

This study has both quantitatively and qualitatively supported the contention that expectations are at the core of any adequate theory of the interest rate structure. Institutional factors play a role but primarily in slowing adjustments and in an indirect way through expectations. Supply comes out as an extremely strong influence tempered only by the wisdom of those issuing securities and their desire to place new issues with a minimum of upset to the existing market. This accounts for the significance and strength of the first two variables in the regression equation and for the failure to find any significant impact from new issues per se. However, it is the demand side which has accounted for most of the change in yield differentials, and it is when the demand for bonds is considered that we must turn away from simple monetary analysis involving changes in available funds to an expectational analysis based upon estimated risk and return. Yield differentials depend upon the allocation of funds among substitutable assets, and a theory which tells us only that the funds will increase under certain circumstances tells us nothing about how these funds will be used. It is for this reason that emphasis has been placed upon interpretation of the last three independent variables as expectational variables. It is conditions in the money market, the capital market and the economy in general that mold expectations, and it is the subjective evaluation of risk and return which serves as the basis for portfolio selection—the ultimate determinant of relative prices and therefore of yield differentials.

APPENDIX

Residuals from Baa Bond Adjustment Regression

Bond #	Dec. 1959		1958		1957		1956		1955		1954		Mean Adj.
	1	2	3	4	5	6	7	8	9	10	11	12	
1											+48	+47	+ 48
2									+26	+62	+72	+75	+ 59
3					- 24	+ 7	+ 1	0	- 8	0	- 4	- 8	- 5
4			+ 59	+ 63	+ 9	+16	- 8	0	+ 5	+15	+11	+ 9	+ 18
5	- 4	- 4	+ 10	+ 52	- 76	-98	-78	+ 3	+25	+33	+29	+21	- 7
6	+ 31	+ 22	+ 22	- 7	- 2	-20	-31	-17	- 6	- 6	-24	-31	- 6
7	+ 29	+ 36	+ 19	+ 10	+ 5	-14	-40	-19	- 9	- 9	-11	-24	- 2
8	+ 46	+ 19	+ 23	- 12	+ 21	+10	+24	+27	+33	+27	+26	+13	+ 21
9	+ 19	+ 16	+ 60	+ 82	+ 26	+22	0	+ 8	+ 4	0	- 2	- 3	+ 19
10	- 81	-151	-191	-197	-139	-64	-35	-63	-74	-78	-61	-61	-100
11	- 41	- 33	+ 2	- 8	+ 5	-30	-12	+28	+36	+41	+40	+35	+ 5
12	- 80	- 22	- 31	- 61	- 39	+12	+16	+12	+21	+22	+25	+26	- 8
13	+ 6	0	- 17	- 37	- 15	+15	- 3	-21	-18	-31	-25	-19	- 14
14	+ 49	+ 33	+ 35	+ 74	+ 38	+17	+ 7	-14	-22	-30	-30	-27	+ 11
15	- 12	- 42	+ 36	- 9	+ 37	+ 3	+ 5	-41	-45	-47	-53	-54	- 25
16	+ 25	+ 59	+ 52	- 57	+ 4	+22	+27	+22	+32	+31	+23	+22	+ 22
17	-112	- 92	- 92	- 39	+ 4	- 6	+22	+30	+33	+24	+14	+ 9	- 17
18	+ 9	+ 26	+ 16	- 4	- 26	-49	- 9	+ 7	+11	+10	+ 9	+23	+ 2
19	+121	+105	+122	+139	+ 78	+36	+81	+59	+54	+47	+56	+66	+ 80
20	- 72	- 77	- 77	- 16	+ 11	- 2	- 8	-30	-38	-41	-44	-55	- 37
21	+ 2	+ 9	+ 10	+ 17	- 13	- 6	-17	-24	-29	-36	-40	-39	- 14
22	+ 27	+ 23	- 5	- 24	- 8	+21	- 6	-10	-10	-11	-12	-10	- 2
23	- 12	+ 10	+ 6	+ 18	+ 57	+52	+41	+16	+11	+ 9	+ 3	+12	+ 19
24	- 12	+ 9	+ 13	+ 31	+ 15	+24	+40	+26	+23	+31	+37	+39	+ 23
25	+ 33	+ 16	+ 2	- 11	+ 34	+29	+33	+31	+24	+27	+25	+28	+ 23
26	- 9	- 16	- 39	- 42	- 10	-43	-26	-20	-24	-21	-31	-26	- 26
27	+ 53	+ 29	+ 28	+ 4	+ 6	+28	+ 3	-11	-35	-25	-34	-16	+ 3
28	+ 10	+ 23	+ 17	+ 36	+ 10	+24	+11	+28	+23	+18	+27	+22	+ 21
29	+ 5	+ 3	+ 9	+ 9	+ 7	+ 1	- 7	-23	-28	-32	-54	-60	- 14
30	+ 43	+ 55	+ 54	+ 68	+ 48	+43	+28	+16	+ 9	+ 7	+13	+13	+ 33
31	- 60	- 53	- 61	- 69	- 67	-43	-44	-30	-32	-20	-25	-21	- 44

Residuals $(Y_C - Y_O)$ are calculated from the regression equation $Y = 5.0258 + .027769X_1 - .000503X_1^2$ with the constant term adjusted for each semi-annual period as shown in Table 1.

APPENDIX

Final Baa Bond Adjustments

	1959		1958		1957		1956		1955		1954	
Bond #	1	2	3	4	5	6	7	8	9	10	11	12
1											+48	+48
2									+59	+59	+59	+59
3					- 24	+ 3	+ 3	+ 3	- 5	- 5	- 5	- 5
4			+ 61	+ 61	+ 4	+ 4	+ 4	+ 4	+10	+10'	+10	+10
5	+ 14	+ 14	+ 14	+ 14	- 84	-84	-84	+22	+22	+22	+22	+22
6	+ 25	+ 25	+ 25	- 15	- 15	-15	-15	-15	-17	-17	-17	-17
7	+ 20	+ 20	+ 20	+ 20	+ 20	-24	-24	-24'	-13	-13	-13	-13
8	+ 19	+ 19	+ 19	+ 19	+ 21	+21	+21	+21	+25	+25	+25	+25
9	+ 44	+ 44	+ 44	+ 44	+ 24	+24	+ 1	+ 1	+ 1	+ 1	+ 1	+ 1
10	-152	-152	-152	-152	-152	-54	-54	-54	-69	-69	-69	-69
11	- 20	- 20	- 20	- 20	- 12	-12	-12	+36	+36	+36	+36	+36
12	- 47	- 47	- 47	- 47	- 47	+13	+13	+13	+24	+24	+24	+24
13	- 12	- 12	- 12	- 12	- 9	+15	- 9	-23	-23	-23	-23	-23
14	+ 48	+ 48	+ 48	+ 48	+ 21	+21	+21	-25	-25	-25	-25	-25
15	- 25	- 25	- 25	- 25	+ 15	+15	+15	-48	-48	-48	-48	-48
16	+ 45	+ 45	+ 45	- 57	+ 19	+19	+19	+19	+27	+27	+27	+27
17	- 84	- 84	- 84	- 84	+ 13	+13	+13	+13	+20	+20	+20	+20
18	+ 17	+ 17	+ 17	- 22	- 22	-22	-22	+12	+12	+12	+12	+12
19	+122	+122	+122	+122	+ 64	+64	+64	+64	+56	+56	+56	+56
20	- 75	- 75	- 75	- 4	- 4	- 4	- 4	-42	-42	-42	-42	-42
21	+ 10	+ 10	+ 10	+ 10	- 15	-15	-15	-15	-36	-36	-36	-36
22	+ 24	+ 24	- 11	- 11	- 11	+24	-11	-11	-11	-11	-11	-11
23	+ 6	+ 6	+ 6	+ 6	+ 42	+42	+42	+42	+ 9	+ 9	+ 9	+ 9
24	+ 10	+ 10	+ 10	+ 10	+ 26	+26	+26	+26	+33	+33	+33	+33
25	+ 10	+ 10	+ 10	+ 10	+ 32	+32	+32	+32	+26	+26	+26	+26
26	- 27	- 27	- 27	- 27	- 25	-25	-25	-25	-26	-26	-26	-26
27	+ 29	+ 29	+ 29	+ 29	+ 12	+12	+12	-24	-24	-24	-24	-24
28	+ 22	+ 22	+ 22	+ 22	+ 18	+18	+18	+18	+23	+23	+23	+23
29	+ 6	+ 6	+ 6	+ 6	+ 6	+ 6	-15	-15	-44	-44	-44	-44
30	+ 55	+ 55	+ 55	+ 55	+ 34	+34	+34	+34	+11	+11	+11	+11
31	- 61	- 61	- 61	- 61	- 46	-46	-46	-46	-25	-25	-25	-25

The Final Baa Bond Adjustments were calculated as the mean of the deviations shown on page 54 for periods exhibiting as nearly as possible stable market ratings. These periods were predominantly three two year periods, as indicated, with variations of the middle period 1956-57 in accordance with the pattern on page 54. For a full discussion of the adjustment refer to the text, pp. 12-18.

Summary: This thesis is a study of the role of the monetary system in Kenya, Uganda, and Tanganyika. On the one hand we explore the possibilities of exercising policy control over the financial institutions; on the other, we examine the interrelationship between monetary variables and expenditure flows in the economy.

Money and capital markets in East Africa are rudimentary; commercial banking is predominantly in the hands of three large international banks, with headquarters in London. As a result, it would be quite difficult for a monetary authority to influence the loan supply function of commercial banks. The possibility is suggested of requiring the banks to deposit funds with the monetary authority equal to a certain percentage of their local lending. If the goal is to encourage the expansion of this lending, interest might be paid on these deposits. As with several other monetary tools discussed, it should be possible to differentiate between various categories of lending.

Bank credit in the area is granted primarily to finance internal trade and imports, although agricultural and industrial lending are also significant. Since these latter categories of lending are quite helpful in the development process, and since in the East African context bank lending is an effective channel for the inflow of funds from overseas, the authorities would probably want to encourage this type of lending at all times. With the possible exception of Tanganyika, the evidence suggests that a restriction of bank lending would have little impact on the level of aggregate domestic demand.

This essay was submitted to Yale University in 1962 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics at Yale University.

Dissertation Committee: Professors Robert Triffin (Chairman), William Fellner, and Reginald Green.

*

I would like to acknowledge with thanks the advice, encouragement and criticism of Professors Robert Triffin and William Fellner. Sincere thanks also go to the Earhart Foundation for generous support during a full year of work on this dissertation.

D. C. M.

MONETARY ANALYSIS IN AN UNDERDEVELOPED ECONOMY: A CASE STUDY OF THREE EAST AFRICAN TERRITORIES

Donald C. Mead

1. Introduction

In any economy the study of the monetary system can conveniently be broken into two aspects. The first concerns the relationship between monetary variables and expenditure flows in the economy, while the second looks at the possibilities of exercising some type of policy control over the monetary variables themselves. Both aspects are necessary if we are to understand the means by which a government authority can influence the level and distribution of expenditures through the monetary system.

In the vast recent literature on the economic growth of low-income countries, relatively little attention has been paid to the monetary and financial systems. Lengthy introductory texts generally pass over monetary questions in a few paragraphs or a few pages. Such detailed studies as have been made deal primarily with the second aspect mentioned above, the problem of controlling the financial system in an environment where financial markets are underdeveloped or nonexistent.

Our study is an attempt to explore both these aspects of the monetary system in the economies of the three East African territories Kenya, Uganda, and Tanganyika. In section 2 we examine the problems of exercising policy control over the financial institutions in East Africa. Besides the standard problems arising from the underdeveloped state of the capital markets, we shall be concerned about the fact that the commercial banks in the area are virtually all international, London-based institutions. In sections 3–5 we attempt to evaluate the impact of monetary variables on expenditure flows in the economy. Section 6 presents a summary of our findings.

2. The Control of the Monetary System

A. The Nature of the Monetary Authority

The monetary system of East Africa is administered by a currency board. In the past this board has operated simply by being willing to buy or sell East African currency in exchange for sterling at the fixed rate of one pound sterling for one East African pound. Sterling funds collected in this way were invested, almost entirely in London. If through the returns to these investments the board's assets grew substantially in excess of its monetary liabilities (in practice, over 110% of money outstanding), the excess was distributed to the "owners" of the board, the East African governments.

The currency board can hardly be thought of as a monetary authority in the normal sense of the term. It exercises no control over any monetary parameter, and in fact has almost no power of discretionary action. The one exception to this lies in its limited power to purchase securities of the East African governments. This means that the board is permitted some fiduciary issue; it seems unlikely, however, that this provision has been used so far as a conscious, countercyclical tool. In the latest report of the

currency board, the substantial figures for purchases and sales of local Treasury Bills are explained as follows:

> In order to provide short-term finance for the governments concerned, the board has agreed to invest in local issues of 35, 63 and 91 day Treasury Bills in Kenya, Uganda and Tanganyika within specified limitations.[1]

At most this might be a method of financing a countercyclical fiscal policy; more probably it is simply a means of smoothing out seasonal fluctuations in tax receipts and government spending. In any case, it is far removed from a countercyclical monetary policy, operating through the financial institutions.

If the governments of the East African territories wish to exercise an influence on the monetary situation of the public either directly or through the financial institutions, one of two developments must take place. The local market for East African Treasury Bills must advance to the stage where the board's present authority to buy and sell these bills can be exercised through the market, thereby affecting asset holdings of individuals or balance sheets of commercial banks; or the board must be given additional powers, to permit it to influence the banking system directly. In following sections we shall examine several such methods of control at some length, to see which seem most feasible in the East African context.

In December 1961 Tanganyika achieved complete independence, and Uganda will follow in the fall of 1962. It has not yet been decided when Kenya will become independent, but it is most unlikely that this should be delayed beyond 1963. There has been much talk of the importance of maintaining the present freedom of trade among the three territories; it does not seem likely, however, that economic and political integration will go much beyond that in the near future. The question of whether there should be one monetary authority or three in the area is one which is likely to be settled on the basis of other considerations, such as the degree of political unity in the territories and the rates at which they achieve their independence. However we feel that, whether there is one authority or three, the economies of the territories are different enough and the likely degree of economic integration small enough so it would be wise for the monetary authorities to be able to differentiate among the territories in their policy decisions. When speaking of the monetary authority in following sections, we assume that, whether there are three offices of one institution or three institutions, they do operate in such a way as to be able to apply different policies in different territories.

B. The Markets for Government Securities

It has often been pointed out that one important restriction on the possibilities of monetary control in underdeveloped areas concerns the narrow and limited capital market. Let us look at the situation in the East African territories to see if this comment seems valid there. Turning first to the

---

1. East African Currency Board, Report for the Year Ended 30th June, 1960 (Kenya: Government Printer, 1961), p. 3.

market for long-term securities, we might ask to what extent the three East African governments have borrowed in the local market. The answer is clearest in the case of Uganda, where long-term local borrowing was nil during the period of our study. In the case of Kenya, local interest payments on the public debt were as follows: [1]

	1955	1956	1957	1958
Paid to households in Kenya	-	0.1	0.1	0.2
Paid to companies in Kenya	-	0.1	0.1	0.1

(in millions of pounds)

In Tanganyika the holdings of long-term debt were distributed as follows, as of June 30, 1958: [2]

	Tanganyika	Kenya	Uganda	East Africa
Persons and non-profit inst.	163	105	15	291
Banks and financial inst.	185	151	-	336
Companies	477	96	5	559
Government	1,150	44	-	1,204
Total, all holders in East Africa				2,420
Total, all holders outside East Africa				12,151
Total, all holders				14,627

(thousands of pounds)

In neither of these territories do we have any estimates of the rates of turnover of holdings of these bonds; with such limited amounts outstanding, however, it is most likely that the market for these securities was quite narrow and limited.

The government of Uganda has taken the lead in the sphere of short-term borrowing; it has been issuing Treasury Bills since February, 1958. Early in 1959 the Economist Intelligence Unit reported that

> the Legislative Council has approved a government demand to increase its borrowing power by issuing from £2 mn. to £5 mn. worth of Treasury Bills, but a recent offer of £0.5 mn. has evoked no interest. It is understood, however, that the currency board will subscribe an initial amount of £1.5 mn. [3]

---

1. East Africa High Commission, East African Statistical Department, Domestic Income and Product in Kenya, 1954-1958 (Nairobi: Government Printer, 1959), p. 89.

2. East Africa High Commission, East African Statistical Department, Public Finance of Tanganyika Territory (Dar Es Salaam: Government Printer, 1959), p. 39. Figures exclude long-term loans from Imperial funds, Barclays Overseas Development Corporation, and the Lint and Seed Marketing Board.

3. Economist Intelligence Unit, Three-Month Economic Review: British East Africa, No. 25 (March 1959), p. 3.

At the end of 1959, the government had issued over £3.5 million in Treasury Bills. By the end of 1960, this figure had fallen to £0.8 million. It is clear that a major share of these bills was in the hands of the currency board, which held nearly £1.6 million in Uganda Treasury Bills on June 30, 1960. The small amount held outside the currency board was probably largely in the hands of firms which held them either to earn a return on temporarily idle funds, or as a politically motivated concession to the government. These facts give us little reason to believe that there was a significant market in Uganda for these securities.

Tanganyika began issuing Treasury Bills soon after Uganda; at the end of fiscal 1960 the currency board held nearly £1.5 million of these securities. As far as we can tell, Kenya has not yet undertaken this means of financing, although it is unlikely that it will delay long, in view of the certain market for a limited number of them with currency board. On the whole, though, it seems clear that any attempt at monetary policy which relies on a well-developed market for either long-term or short-term government securities is doomed to failure.

C. The Nature of the Commercial Banks

Having examined briefly the monetary authority and the financial environment in which it must operate, let us move on to the institutions which the authority would seek to control, the commercial banks. We need to know which banks these are, what types they are, and how they determine their lending policies.

The banking business in East Africa is concentrated primarily in the hands of three large banks: Barclays DCO, the Standard Bank of South Africa, and National and Grindlays Bank ltd. There are several other banks operating in East Africa; unfortunately no exact measure of their relative importance is available, since all of them operate in other areas as well, and publish only aggregated balance sheets for their world-wide operations; this makes it impossible to compare deposits in each of the three banks to total deposits in the country, for example. However, it seems reasonable on two counts to concentrate our attention on these "big three"; in the first place, they account for the great majority of the banking offices in East Africa (193 out of 213 in 1960, counting branches, agencies, and sub-branches);[1] secondly, the other banks are relatively recent arrivals. As late as 1950, for example, the Banque du Congo Belge had one branch in Dar es Salaam, which has since been closed; all other banking was at that time in the hands of the big three. In 1955 only three new banks with eight branches had entered the scene. Thus it seems clear that the other six banks now operating there, with their 20 offices, are relatively unimportant compared to the big three.

---

1. These figures, as well as those on the distribution of branches of the individual banks given below, are taken from The Bankers' Almanac and Yearbook: 1959-1960 (New York: Thomas Skinner and Company, 1960).

These organizations all are large international banks, with head offices in London and branches around the globe. Their East African business forms but a small part of their immense operations; this can be seen from the fact that, at the end of 1959, the total of their world-wide assets came to £1,127 million, while the total assets of the whole banking system in East Africa was only £150 million at the same date.

As we mentioned, none of these banks publish statistics which would permit us to separate out balance sheets country by country; therefore it is impossible to be completely precise about how their business is distributed geographically over the globe. However, we do have an indicator which permits us to be fairly accurate in describing the world-wide pattern of their business; this is the distribution of banking offices. Of course this will not be completely accurate; one large office might do the same business as a number of small ones. However, the fact that in 1959 the Standard Bank of South Africa had its 815 offices distributed with 660 in South Africa, 89 in the Central African Federation, and 56 in East Africa permits us to get a fair picture of the geographical distribution of its business. The corresponding pattern for Barclays DCO is as follows: out of a total of 1,175 offices, 670 were in South Africa, 96 in Central Africa, 128 in West Africa, 62 in the West Indies, 95 in East Africa, and 124 scattered through 12 other countries. Here again it would seem that it is in South Africa that the major part of the business is done, although in this case there is a much wider dispersion among the other areas. In the case of the third bank, National and Grindlays, there is a shift from Africa to Asia; the distribution of branches at the end of 1958 was as follows: 18 in India, 12 in Pakistan, 20 in five other countries of Asia and Africa, and 44 in East Africa.

We must now address ourselves to the question of the determination of the lending policies of the East African branches of these banks. From the point of view of the possibilities of monetary policy, the question is whether or not they vary their lending policies in response to changes in any variable which is within the control of a local monetary authority.

This question might be rephrased by asking whether it is reasonable to expect banks to have a varying supply curve of loans in East Africa, and if so, what variables would cause it to fluctuate. Posing the question in this way, we must make clear what we mean by the term supply curve or supply function of loans. In a country with a well-developed financial system, the term refers to the functional relationship between the amount of loans outstanding and the interest rate, showing that at higher rates banks are willing to lend more. In underdeveloped countries, where money markets are imperfect and interest rates relatively insensitive to market forces, it would be better to speak of the amount of loans which the banks would be willing to make in the face of a given demand situation. We might think of an array of loan petitioners with projects of increasing riskiness, perhaps with the interest charge to all approximately equal; an outward shift of the supply schedule would then mean that loans are made to more risky applicants. If these more risky loans were made at a higher interest rate, that fact would be irrelevant; the point is that, before the schedule shift, the additional loans would not have been made at all.

In the literature it has remained an unanswered question how these supply considerations have influenced bank lending patterns in colonial areas. Greaves reports that her interviews with bank managers convinced her that standards of "bankable loans" were fixed and unchanging; it was London's job to worry about reserves, while overseas managers could grant all loans that came up to a certain fixed standard, and no others.[1] If this is the case, then the supply schedule is clearly constant, and a monetary policy which aimed at shifting the schedule would be fruitless.

Newlyn and Rowan are not willing to go this far.[2] They simply indicate that 1) banks are not strongly influenced by their local reserve position, and 2) local developments in one area (e.g. an export surplus, or an increased demand for loans) are likely to have but small effect on the overall reserve position and loan policy of the banks as a whole.

We would never expect a bank to try to keep its liquidity ratio at 32%, or 36%, or any other figure, under all circumstances. The existence of a changing supply curve would simply mean that a bank would look differently at a loan application when its liquidity ratio was 45% than when it was 35%. The question we must examine is whether or not there is in fact any such liquidity effect on bank lending, and if so, if the relevant liquidity ratio concerns the bank's world-wide balance sheet, or that within one country or some combination of these two.

It is quite difficult to arrive at a quantitative answer to this question, both because of the number of variables involved (particularly with regard to the demand for loans) and due to the degree of aggregation of the data. However, if the answer is difficult to attain, it is also extremely important from the point of view of the methods and possibilities of monetary policy in the area. If the commercial banks there are concerned only about their world-wide portfolio, then monetary policy using standard tools within a segment as small as East Africa would be extraordinarily difficult. For example, the immense open market operations or variations in reserve requirements which would be necessary to affect the world-wide balance sheet of Barclays DCO would be unthinkable for an East African monetary authority alone.

There are two sets of evidence which help us to make a judgment on this question. The first is the bankers' testimony before the Radcliffe Committee, while the second concerns the variations in the liquidity ratios which the banks have permitted in the past.

The testimony before the Committee on the Workings of the Monetary System (Radcliffe Committee) provides us with a rich source of information about the operating rules of British overseas banks, some of which is particularly relevant to the present question. For example, Professor Cairncross

---

1. Ida Greaves, Colonial Monetary Conditions ("Colonial Research Studies," No. 10; London: Her Majesty's Stationery Office, 1953), pp. 46-47.

2. W. T. Newlyn and D. C. Rowan, Money and Banking in British Colonial Africa (Oxford: The Clarendon Press, 1954), pp. 159-60 and 162-63.

asked the following of Mr. R. D. Smith, the General Manager of Barclays DCO:

> Professor Cairncross: Your liquidity cannot be equal in all directions all round. There must be some places where you are, from a local point of view, in a rather liquid state. Do you apply the same credit price throughout? Mr. Smith: Yes. One of the advantages of banking in various territories as we do is that one can meet seasonal demands in one territory by utilizing some of the larger credit balances which are maintained seasonally in other territories.[1]

Another discussion between Cairncross, Smith and Mr. Wilson, London Manager of the Standard Bank of South Africa, runs as follows:

> Professor Cairncross: I assume that there are some territories in Africa where you lend by drawing, in effect, on deposits made in other parts of the continent? Mr. Smith: Yes.
>
> Professor Cairncross: Is this mainly in East Africa? Mr. Wilson: Speaking for my own bank, I would say that in South Africa, also, before the credit squeeze we had no lack of opportunity for making good banking advances.
>
> Professor Cairncross: It is rather important to know how far you are prepared in your business to make advances in one country substantially in excess of deposits in that country? Mr. Smith: One would normally only do that as a very temporary measure. We do tend to meet seasonal requirements.[2]

This discussion would make it seem likely that, for temporary seasonal needs, banks pay little attention to local liquidity ratios. When it comes to more lasting patterns, on the other hand, they seem to be more concerned about keeping some balance between loans and deposits for example. They allow their national liquidity ratios to fluctuate much more freely than if the branches within one country were independent institutions; but there also seem to be fairly strong limits to these fluctuations, for any but temporary seasonal needs.

The reaction of the banking system to the financial developments of 1960 give us a further insight into the determinants of bank lending in East Africa. During the course of that year, because of the large import surplus in Kenya and the large-scale capital flight due to the uncertainty over the future stability of the area, the net claims of the banking system on the rest of the world fell by over £18 million. This brought about a sharp deterioration in bank liquidity, which "finally forced the banks to take positive steps to restrict further expansion of credit."[3] Interest charges were increased,

1. Great Britain, Committee on the Working of the Monetary System, Minutes of Evidence (London: Her Majesty's Stationery Office, 1960), par. 4367.

2. Ibid., pars. 4337-4376.

3. Economist Intelligence Unit, Three-Month Economic Review: British East Africa, No. 32 (December 1960), pp. 4-5.

and an over-all attempt was made to reduce the level of bank lending.

On the whole, it seems clear that the international character of the banks permits them to be less bound by national liquidity considerations than independent organizations without the substantial reserves of overseas head offices to fall back on; but the available pieces of evidence seem to support the hypothesis that even these international banks do take account of their national portfolio and liquidity situation, varying their lending policies accordingly. If the range of tolerance is wider for them, this would mean simply that a monetary policy would have to be pushed more vigorously before it would influence the level of bank lending, not that this type of policy control would be impossible in that situation.

There are certain possible policy moves which might induce the banks to pay more attention to their portfolio situation within each country. We mention two: controlling capital movements, and reducing the confidence of the banks in exchange rate stability. These would make it undesirable or impossible for the banks to count on foreign assets as backing for local loans; the banks would recognize the possibility that, when these foreign assets were needed locally, they could not be brought in freely; or alternatively, when the assets were no longer needed locally, they could not be freely transferred abroad, or could be taken out only at a less favorable exchange rate. The result should be to increase the degree to which the international banks would treat their operations in each country as separate units, trying to keep some desirable portfolio balance within each country. The great disadvantage of this type of policy is that it would also bring about the undesirable result of slowing down the inflow of private capital for investment and development purposes. Even within the banking system itself, it is likely that the result would be to gain a measure of control over bank lending at the cost of lowering the general level of this lending. For this reason, it seems not to be a fruitful approach.

In conclusion, we feel that it is reasonable to assume that the large, international banks in East Africa are willing to permit fairly substantial fluctuations in their domestic liquidity ratio without taking corrective action; beyond a certain range, however, changes in their liquidity ratio would bring about an alteration in their loan supply function. Let us see what this implies about various methods of monetary control which the authorities might adopt there.

D. Methods of Control

Perhaps we might start with the most direct approach to the control of bank lending, which would be for the government to set a maximum on the value of loans which any bank could make in a given period. Such a ceiling might be imposed on specific types of loans which the authorities wished to control, such as urban mortgage lending. This is an asymmetrical tool; the monetary authority could use it to restrain lending, but not to encourage it, except in the negative sense of removing the restraint. If the monetary authority expects that the commercial banks might lend more during a given

period than is "good" for the economy, then such a method of direct control might be an effective way of cutting bank lending to the "right" level.

Besides the asymmetry already mentioned, this approach is subject to two further difficulties. The first is in knowing the level of lending which is "good" for the economy. This kind of judgment is implicit in any monetary policy; yet the fact that, in this approach, the decision must be made explicit, without any possibility of flexibility through the market, makes it a more difficult tool to use than more indirect methods. Furthermore, once the decision is made as to the over-all level of lending to be permitted during the period, the question arises as to the most equitable way of allocating these loans among various banks. The simplest approach would be to limit the increase in lending of each bank in a given period to some percentage of its loans outstanding at the beginning of the period. The problem with any such approach is that it would tend to freeze the existing pattern of importance among the various banks; in a period of rising nationalism; and with the leading banks being all London-based, this is bound to bring cries of discrimination. By setting the permitted increase in both value terms and percentage of starting lending, letting the banks use the larger of the two, one might reduce this difficulty, although even this would not be entirely satisfactory.

For all these reasons, it does not seem likely that this type of direct control of bank lending would play a major role in the "box of tools" of a monetary authority in East Africa. However, if the problems faced appear to be of a specific type (i.e. a periodic or permanent tendency toward excessive bank credit, either in the aggregate or for specific purposes) and if other means of control do not work satisfactorily, then surely this approach could be used.

Indirect methods of control over bank lending are those which influence the banks' willingness to lend (i.e. their supply function of loans) as well as the demand for loans by borrowers. In principle, the monetary authority might try to influence a bank's supply function through two closely connected channels: by controlling the rate of return which the banks earn on certain alternative securities, or by changing the bank's liquidity situation. If the first of these spills over to the interest charge on bank loans, it might have an impact on the demand for loans as well. Let us look at each of these aspects in turn.

Taking first the influence of interest rate changes on the level of lending, this influence might operate in three ways. In the first place, changes in interest rates bring about concomitant capital gains and losses, particularly on longer-term securities. As developed within the context of the United States, this "availability approach" depends primarily on the effect of these capital gains and losses on the willingness of banks to lend to private borrowers.[1] For it to be a workable tool of monetary control, banks must have among their assets a substantial volume of such long-term securities, the rate of return

1. For a full exposition of this approach, see Robert V. Roosa, "Interest Rates and the Central Bank," Money, Trade and Economic Growth, Essays in Honor of John Henry Williams (New York: The MacMillan Company, 1951), pp. 270-95.

on which would be subject to control by the monetary authority. This is clearly not the case in East Africa.

A second possible role of the interest rate as a determinant of the level of bank lending would be through its influence on the demand for loans. It is fairly widely agreed, however, that this factor is only significant in the case of such long-term projects as houses and public utilities. Since bank lending for such projects was small in East Africa, we feel justified in neglecting this effect. [1]

There is still a third possibility of interest rates influencing the level of lending; this would be the case where tradition, competition, or institutional reasons deter the banks from changing their charges on loans to the public, even over the course of shifts in the rate of return on alternative assets. If this were so, then a rise in the Treasury Bill rate might make the banks less ready to supply funds at the going commercial loan rate to support a new private project.

Since there are no complete published data on the rates of interest charged on commercial loans in East Africa, it is difficult to judge the importance of this effect. Such scattered information as we have, however, as well as the discussion of interest rate determination in the Radcliffe testimony, indicates that the banks do vary their charges to private borrowers in accordance with general credit conditions. We see no reason to believe that this type of influence of interest rates on the supply of loans was a significant factor.

The other main area of indirect monetary control includes those methods which attempt to influence the level of bank lending through changes in the liquidity of the banks' asset portfolios. The best example of this approach is the establishment of some type of reserve requirement. This usually takes the form of a requirement that the banks deposit with the monetary authority a certain percentage of their demand deposit liabilities. By varying the proportion of the banks' assets which are immobilized in this way, the authority can have a substantial influence on the banks' liquidity situation, and thus on their willingness to lend.

This approach would be more difficult for East Africa because of the international character of the commercial banks there; as we have seen, these banks are probably willing to accept wider changes in their liquidity situation without any induced change in their loan supply function. On the other hand, as we indicated above, this does not necessarily mean that a monetary policy of this type would be impossible there; rather it means that the authority would have to push any given policy harder before it would affect the banks' willingness to lend.

It is possible to introduce a variation into this general approach which would make it more effective in the East African context; this would be to make the required deposits with the central bank a function of bank loans

1. See also Committee on the Working of the Monetary System, op. cit., pars. 4425-4428, where the overseas bankers indicate that "putting up the interest rate does not really deter borrowers."

outstanding, rather than of deposits. By striking directly at the profitability of commercial loans, the authority might be more effective in influencing the level of lending than if it tried to work through the over-all portfolio balance. It would also be possible, using this type of approach, to have different deposit requirements depending on the purposes for which the loans were made. This would permit the authorities to be selective among the various types of loans in their credit tightness.

If this approach were used, then admittedly the notion of deposits with the authorities to serve as reserves—as funds which would be drawn upon in an emergency—is gone. This type of purpose for the special deposits could be restored by giving the monetary authority the option of freeing a portion of these reserves in the case of a general withdrawal of deposits. It seems to us, however, that the bankers in East Africa have been wise and foresighted enough to do without this type of reserve for a number of years; what is needed now is not a requirement which would make deposits more safe, but rather one which would permit the authorities some measure of control over private bank lending. This the suggested approach seems likely to do.

There would be a simple way of converting this technique of restrictive monetary control into one for inducing expansion; this would be to pay interest on the deposits with the central bank. In times when expansion was desired, if this rate of interest were set at (say) one percent higher than the going rate on Treasury Bills, this might encourage the banks to be more active in seeking outlets for loans, thereby earning the right to a higher level of high-interest-earning deposits.[1] In this situation the deposit of such interest-bearing reserves with the central bank should be thought of as a privilege rather than a requirement; the goal would be to make private lending more attractive to the commercial banks. The central bank could reloan funds received in this way to the government, thereby permitting the government in turn to follow the same expansionary line in its fiscal policy. It would still be possible to pay different rates on different classes of loans, thereby encouraging some types of loans while discouraging others. This general approach seems to have quite interesting possibilities.

The variation of the bank rate as a tool of monetary policy falls between influences that work through interest rates and those which operate on banks' liquidity; for it to be effective, it must combine an illiquid situation in the banks' portfolio with either a high bank rate (to discourage borrowing from the central bank) or a low one (to encourage it). The possibilities of using this tool as a means of restricting bank lending in East Africa are sharply limited by the international character of the banks; if the central bank borrowing rates move too high there, the banks could simply draw on their head offices overseas, or (if necessary) borrow elsewhere.

An expansionary bank rate policy, whereby the central bank agreed to

1. In order to avoid charges of discrimination in favor of large banks and to keep cost low, this bonus could be restricted to deposits made in conjunction with new loans issued after the expansionary policy was announced.

discount securities at low rates, might be somewhat more successful; once again it is the nature of the commercial banks which would limit its effectiveness. If the commercial banks were relatively illiquid on their world-wide balance sheets as well as in East Africa, then the privilege of discounting at low charges in East Africa might be used as an opportunity for the bank to increase its world-wide liquidity, with little impact on local lending; alternatively, any induced increase in lending might be spread over the world-wide network of branches, leaving a negligible effect in East Africa. This instrument is likely to be more effective in the situation where the commercial banks are reasonably liquid on their world-wide balance sheet but illiquid in East Africa taken separately; then a willingness on the part of the central bank to discount securities might allow the banks to increase their lending in East Africa. This tool cannot cause the banks to want to lend more; it can only permit them to increase their lending, if they so desire.

To be effective, such a technique would have to differ from that practiced in Great Britain in two respects. In the first place, in the absence of financial intermediaries to play the role of the discount houses in the British market, the commercial banks themselves would have to be willing to borrow from the central bank. There is nothing strange about this to American ears, but it might not find as quick acceptance among bankers brought up on British traditions. Secondly, if the banks hold only small amounts of government bills and securities, then in order to be effective the discounting would have to be done for other bank assets as well. This should not present any insuperable obstacles, although traditionalists might object. On the whole, bank rate policy might prove quite useful as an expansionary tool.

Along with the bank rate, there is another tool of monetary policy which works both through interest rate changes and through the banks' overall portfolio liquidity; this is the technique of open market operations. We feel that the extremely limited financial market virtually rules out this approach in East Africa for the present and in the foreseeable future.

Perhaps we should mention here the possibility of moral suasion. This seems to have played a considerable role in the Central African Federation, [1] and could well do so in East Africa as well. It has been pointed out that the use of this tool is not contingent on the establishment of a central bank; the government or the currency board could try their hand at it now. [2] It seems, however, that it is more likely to be effective in the hands of an institution which has the authority to enforce its will if moral suasion fails.

E. Other Financial Institutions.

In recent years there has been a proliferation of these institutions in

1. Committee on the Working of the Monetary System, op. cit., par. 4355.

2. International Bank for Reconstruction and Development, The Economic Development of Tanganyika (Baltimore: Johns Hopkins Press, 1961), pp. 336-37.

East Africa. Broadly speaking, their functions can be separated into two categories. The first is to receive deposits, or in other ways to encourage the habit of saving on the part of the public; this function is exemplified by the post office savings banks and the insurance companies. Secondly, they are means for directing available funds into specific types of organization. In many cases both of these aspects are present to some extent; the Uganda Land Bank, whose primary purpose is to make agricultural loans, also accepts deposits, and the insurance companies have made some loans in East Africa; but conceptually—and in most cases, practically—the two functions can be kept separate. The funds which these organizations receive from the public but do not need for local lending purposes are normally invested in London.

We might discuss the importance of these two organizations from two points of view: what their position has been in the economy up to now, and what developments we might expect for the future, particularly taking into account the possibilities of policy control.

Looking first at their role in the economy to date, in most cases there are no data to indicate the precise breakdown between funds in London and local investment of the various organizations. The indications are that, for most of them, the amount loaned out in East Africa was rather small. The main exception to this seems to have been the life insurance companies; for that group as a whole, total income in East Africa from sources other than premiums came to over £550,000 in 1958.[2] On the presumption that the largest part of this total came from interest on local loans, it seems that lending by these companies in East Africa was quite substantial. For the other types of insurance companies, such income was negligible.

With regard to the land banks, these were only of real importance in Kenya, where the balance of loans outstanding reached £3 million in 1959. By that date lending was just over a million pounds in Tanganyika, and one third of a million in Uganda. The territorial post office savings banks had deposit liabilities of nearly £13 million at the end of 1959. The following table gives a breakdown of number of accounts in these banks in Uganda. Unfortunately this breakdown is not available by value of accounts; the table does indicate that substantial numbers of Africans are using the banks, even if the amount each has to save is still small. In the absence of any published details on the asset holdings of the savings banks, we presume that they have followed the pattern of other British dependencies and invested by far the greater part of their funds in London.

---

1. International Bank for Reconstruction and Development, The Economic Development of Tanganyika (Baltimore: Johns Hopkins Press, 1961), pp. 336-37.

2. East Africa High Commission, East African Statistical Department, East African Insurance Statistics, 1950-1958 (Nairobi: Government Printer, 1961), p. 18.

	Number of accounts in Uganda Post Office Savings Banks, as of Dec. 31, 1960
European	2,185
Asian	13,004
African	107,149
Trustee	1,675

Source: Great Britain, Colonial Office, Uganda: Report for the Year 1960 (London: Her Majesty's Stationery Office, 1961), p. 38.

With regard to the future, it is clear that the primary question concerns the extent to which these organizations can be encouraged or induced to increase their lending in East Africa. One important development in this regard is the creation of a market for Treasury Bills of the three governments; another is the progress being made in the establishment of full and transferable land rights, thereby facilitating the development of mortgage lending. The indications are that these financial institutions could become quite important suppliers of capital in years ahead.

If this is correct, then variations in the rate of new lending by these organizations might come to have an important impact on the monetary situation in East Africa. Direct policy control of semi-governmental groups such as the land banks and the post office savings banks with a view to some type of monetary stabilization might be fairly easy to institute, and might develop into an important means of influencing the level of monetary assets outstanding.

We should recognize, however, that these institutions concentrate their lending primarily in areas which are directly related to the process of economic development, such as agricultural improvement and industrial expansion; furthermore they are often willing to make longer-term loans than the commercial banks could safely undertake, thereby filling an important gap in the financial structure of these economies. As a result, the authorities would be quite hesitant about restricting this type of lending in an attempt to influence the over-all monetary situation. A more reasonable policy goal would be to seek to achieve a continuing expansion of lending by these institutions; if monetary restraint is needed, it could better be sought through other channels less directly connected with the process of development, such as commercial bank lending to finance internal trading and imports. In sum, it seems that monetary control through these other financial institutions would be relatively easy to achieve, but probably would not be wise in view of the crucial importance of the type of lending which they provide in the course of economic development.

3. The Analytical Framework

In discussing the impact of monetary variables on expenditure flows in the economy, we need a framework in which to place the analysis. A

simple Keynesian model seems satisfactory for our needs, in spite of numerous criticisms which have been leveled at it as a framework in which to examine economic development as a whole. We shall start by presenting a simplified version of that model, in which monetary parameters play no role; we then introduce the monetary variables into the system in such a way as to be able to test their importance, using available data for the East African territories.

A. The Model with a Neutral Monetary System

The concept of a neutral monetary system has been used by various people in various ways; for our purposes, it refers to a system under which induced changes in the level of financial assets outstanding can be ignored in a discussion of income and expenditure. We hope that this approach will give some indication of the assumptions we should introduce into the model concerning the monetary system, in the following section, to bring us closer to reality. Let us begin by specifying the major sectors of the economy, as follows:

1) The subsistence sector, S. This sector could be treated in either of two ways. It could be taken as completely independent of forces in the rest of the economy, either increasing through time (through better technology or use of more resources) or decreasing through time (as more of the given product is brought to the market); or alternatively it could be regarded as being influenced by the state of the rest of the economy (e.g. falling as local demand rises, or rising when world prices fall). The nature of this function really depends on the nature of the subsistence production and its responsiveness to changes in the market sector. Unreliable data and short periods make it impossible to discern any clear patterns in this regard in East Africa. The figures available indicate a nearly constant or slowly rising level of subsistence production during the 1950's; this generally resulted in some reduction in the share of this sector in total product as the rest of the economy grew. Year-to-year changes seem not to have been influenced by developments in the market sector. We shall assume that the two parts of the economy were entirely independent; in the remainder of this study we shall leave the subsistence sector aside entirely, concentrating our attention on monetary income and product.

2) The export sector, E. This consists of all domestic product sold abroad. In many cases this is the most widely fluctuating component of domestic income and product. Because the income from exports (particularly in these primary producing countries) is so heavily determined by world demand and local weather conditions, we shall consider it to be exogenously given. This assumption is not entirely correct; farmers do vary their plantings in response to market forces. As an approximation, however, we shall take export earnings as given.

3) Domestic production for domestic markets, H (for home). This is probably the most crucial sector in the course of economic development. Its responsiveness to demand shifts is of particular concern in avoiding inflationary pressures, and depletion of international reserves. The total is made up of two components: consumer goods (C_h) and investment (I_h).

The demand for consumer goods might best be thought of as a function of Y; in simplest form, we could write $C_h = aY + b$. With regard to domestic capital formation, for the moment we shall take it to be autonomously given ($I_h = \bar{I}_h$). Due to the slowness of response on the supply side to changes in the level of demand, we might also indicate that there is a restraint on the possibility of expanding production in this sector this period, which we might call H^t, beyond that of the previous period, H^{t-1}. This inflexibility of supply could best be explained by saying that, although the needed factors of production are available in the economy, the markets for these factors operate rather sluggishly; when there is a new call for factors to produce some good for which demand has increased, the factors do come forward for employment, but only gradually. This might be expressed as a side-condition, such as

$$C_h^t + I_h^t = H^t \leq 1.05\, H^{t-1}.$$

We should recognize the further category of spending on imports, M. This is not a sector of production, as the others are, but rather of expenditure. Imports are also made up of consumer goods (C_m) and investment goods (I_m).

The equality of income and allocation of product in this world is expressed as

$$H + E = H + M,$$

implying that $E = M$. If foreign trade is not balanced, income is less than product absorbed by the amount of the deficit, D (or greater than expenditure by $-D$); in general,

$$H + E \equiv H + M - D' \equiv Y \text{ (money national income).}$$

As indicated, these are identities.

The level of income in any period is now uniquely determined. Using the simplified equations given above, we can write

$$Y = H + E = aY + b + \bar{I}_h + \bar{E}$$
$$= (b + \bar{I}_h + \bar{E})/(1-a),$$

where $1/1-a$ is related to the standard Keynesian "investment multiplier."[1]

The main criticism which has been leveled at this simple multiplier-type model concerns supply elasticities. The model was orginally designed to look at cases of less than full employment, when aggregate supply could and would expand in response to an increase in aggregate demand. If this is true only to a limited extent, then the marginal propensity to consume will be

1. It is not quite equal to it, in its usual formulation, since a is the marginal propensity to consume home-produced goods. At this stage this seems simpler than including total marginal consumption in a and then subtracting out marginal imports, as we shall do on p. 78 below.

a function of the extent to which the supply of goods will rise to meet new demands.

The boundary condition on output changes which we included in our discussion above might be thought of as an attempt to allow for this difficulty; it is not entirely satisfactory in this regard, however, since it leaves many of the most difficult questions unanswered. For example, it tells us little about the relevant time dimension, nor about what happens when the desired level of expansion exceeds the indicated attainable one. For these reasons, we feel that there are rather sharp limits on the extent to which this type of model could give quantitative estimates of new equilibrium points after an autonomous change, for example. However, we do feel that the model is still useful to give a qualitative indication of forces at work and directions of movement. Furthermore, there does not seem to be any very satisfactory alternative framework in which to place our examination. In short, we feel that this model is the best available framework for our analysis, and can be quite informative if it is not pushed too far.

Having set forth this simplified model, then, our intention is to see what it implies about the monetary aspects of the economy, and to proceed by introducing assumptions into the model which would take account of these monetary developments.

Going back to a situation of equilibrium, in the sense that expenditure in all categories by all income recipients is fully adjusted to income, let us first of all make clear that it is not necessarily true that foreign trade will be balanced; in fact it would only be by chance that this would occur. Taking the country as a whole, residents can spend more than their income only by importing more than they export; this they can do only by drawing down their monetary assets. Unless offset by fiscal policy, by bank or foreign lending, the money supply or other monetary assets must fall by an amount just equal to the deficit, and this fall will be repeated each period for as long as the deficit continues. Similarly if the people of the country save more than they invest, which they can do (ex post) only by exporting more than they import, then the monetary counterpart of this is an equal increase in the supply of monetary assets of the public. In like manner, this rise in assets will be repeated each period, for as long as the country remains in surplus.

The monetary assumptions underlying the model can now be made explicit. With regard to individual income recipients, the assumption is that their spending propensities are unaffected by their stock of monetary assets; the demand for money is infinitely elastic with respect to income. This assumption is particularly difficult to accept in times of a deficit, where (still assuming away banks and foreign lending) the money supply would dwindle toward zero, apparently without any effect on either income or expenditure. It is the implication that we must push aside these changes in the money supply and focus only on income as the determinant of expenditures, even over substantial shifts in the money supply, which points up the inadequacy of this simple model.

If we try to introduce a banking system to remedy this situation, we

find that we need two rather special assumptions, one concerning the demand for loans, the second the supply of loans. Taking the latter first, the necessary assumption is that the banks' willingness to lend would be unaffected by their liquidity situation, at least as determined by the local balance of payments situation. This means that, in times of foreign trade surplus, banks would continue to accumulate deposits and liquid assets, without any induced change in their local lending policies; in times of a deficit, on the other hand, they would be willing to increase their loans as much as needed, drawing on overseas reserves when necessary, to finance excess imports.

The shortage of foreign exchange aspect of this situation as a force leading to change of government policy has been widely recognized; yet in countries with no central bank and in which the private banking system consists primarily of branches of overseas banks, the situation is different. There the using up of foreign exchange would not show up in the figures of an exchange equalization account, but in a fall in the domestic money supply, or else in a shift of the national loan-deposit ratio of the commercial banks. If there is any induced policy change caused by an increased demand for foreign exchange, it would be in the loan policies of the banks rather than in any government agency. The assumption of monetary neutrality implies that there is no such induced policy change on the part of the banks, that their willingness to lend is unaffected by their local liquidity situation.

In section two we discussed at some length the factors which determine the banks' lending policies in East Africa. As we indicated there, the fact that the commercial banks are predominantly large international organizations gives them greater freedom to permit variations in their liquidity ratios, although this flexibility is not without limits. Our present concern, however, is with the reactions of individual income recipients to changes in their monetary asset holdings, assuming that there is no induced change in the commercial banks' lending policies. From this point of view, the assumption of monetary neutrality implies that, over whatever period the model is being used, the propensity to spend is not affected by changes in the level of net monetary assets of the public. That is, people would be willing to repay loans and to accumulate balances continuously (in the case of a trade surplus), or else to borrow and keep on borrowing if they have to (as they will with a continuing deficit), to finance any level of expenditure which their (unchanging) income and preferences indicate; neither borrowing nor accumulation would affect their propensity to spend, and hence the equilibrium income level. With the exception of the willingness to borrow whatever is "necessary" to finance the given expenditure level, we cannot say anything about the demand for money.

Perhaps we should emphasize that this conclusion is not meant to be a description of actual events; on the contrary, it is far removed from reality. It is designed to give specific meaning to what we have referred to as a neutral monetary system. Our next task is to introduce monetary parameters into the model in such a way as to indicate how these implied monetary developments in the neutral system would reflect back on expenditure flows in the economy, or would be tempered by changes in these flows induced by the monetary developments themselves.

B. Monetary Parameters Introduced into the Model

Our intention now is to set forth an empirically verifiable hypothesis concerning the role of the monetary variables in influencing the level of consumption spending in the economy. The interaction between monetary variables and the level of investment will be examined in section 5; in the meantime we shall take the level of investment spending as given. Furthermore, to start out this section, let us examine a world without banks or government, with foreign trade in balance, and with stable prices, so that values can be taken to represent either real or monetary magnitudes. All of these assumptions will be removed in the course of the discussion of this section.

In examining the relationship between monetary assets and consumption spending, it is tempting to start from a functional relationship between the level of income and the demand for money; one might then argue that an excess of actual money holdings over the desired level (as indicated by the functional relationship) would cause spending to rise in an attempt to draw down the excess money holdings. This approach opens up a whole host of new problems and difficulties, however; it therefore seems wisest to avoid this detour entirely, simply hypothesizing that the level of consumption spending is influenced by the size of the consumer's money holdings as well as by his income. In simplest terms, this might be expressed as:

$$C = cY + bMo + k .$$

In this formulation, c is the marginal propensity to consume, and is assumed to be constant; b indicates the impact on consumption spending of the level of money holdings Mo; it is also assumed to be constant. It could be thought of as a measure of the shift in the consumption function as a whole due to monetary changes, leaving the slope coefficient c unchanged.

The meaning of this formulation might become clearer if we divide the equation through by Y, giving

$$C/Y = c + bMo/Y + k/Y .$$

That is, the average propensity to consume is a function of Mo/Y and Y.

We could simplify our hypothesis substantially if we assume that $k = 0$. In a model in which money did not enter the consumption function, this would imply that the average propensity to consume is constant, and equal to c. In the present model, setting $k = 0$ means that the average propensity to consume will be constant if Mo/Y is constant, although the average propensity to consume will not be equal to c unless b also equals zero.[1]

---

1. We shall see below that, while data on this point are particularly weak, it seems reasonable to regard income per capita as approximately constant during the period of our study. In this situation, setting $k = 0$ is less likely to lead to a distortion than if income per capita were fluctuating more widely.

On this basis, and using the assumptions specified in the beginning of this section, we can complete the model as follows:

$$I = I_m + I_h = \bar{I}$$

$$NFI = E - M = O$$

$$Y = C + \bar{I}$$

$$= (bMo + \bar{I})/1-c).$$

The standard investment multiplier in this model would be

$$dY/dI = 1/1-c.$$

If instead of this we took the derivative of Y with respect to Mo, we would get

$$dY/dMo = b/1-c.$$

Following the pattern above, this might be called the money multiplier, since it tells the change in equilibrium income associated with a given change in the money supply. This terminology is somewhat misleading; since we have taken the derivative of Y with respect to a stock rather than a flow, the time dimension will be different.[1] In fact it is a marginal velocity concept, the reciprocal of the marginal "Cambridge k." Formulating it in the way we have permits us to say something further about why this money multiplier works the way it does. The value of b tells the initial increase in consumption spending caused by the rise in Mo; the value of $1/1-c$ tells the extent to which this increase in income is respent on domestic products, thereby inducing a multiple increase in income. We might say that b tells the extent of the original upward shift of the consumption function, while $1 - c$ tells the extent of the movement along the function.

Perhaps an example would help to clarify our approach. We still assume that we are in a country without banks or government; foreign trade is originally balanced, and the marginal propensity to import is zero. Investment $\bar{I} = 100$; the money supply $Mo = 200$. Suppose that the parameters of the consumption function can be specified as follows:

$$c = 0.7;\ b = 0.2;\ k = 0.$$

That is, $C = 0.7\ Y + 0.2\ Mo$

We can now solve our model as follows:

$$Y = (bMo + I)/1-c$$

$$= [(0.2)(200) + 100]/1-0.7$$

$$= 140/0.3 = 466.7$$

---

1. The result will be in units per year, rather than a pure number, as the investment multiplier is.

Now suppose that world demand for the country's product rises sharply for one period, so that export earnings increase by 100. Income rises by some multiple of this amount; since the marginal propensity to import is assumed to be zero, and there are no banks, the money supply will rise by an amount equal to the trade surplus, or 100. If in the following period world demand falls back to its original level, then all parameters would return to their original values except the money supply, which would remain at 300. The result is that, in the following period,

$$Y = [(0.2)(300) + 100]/1-0.7$$
$$= 160/0.3 = 533.3$$

This same result could also have been obtained from the equation at the top of page 76:

$$dY/dMo = b/1-c$$
$$= 0.2/1-0.7$$
$$= 0.667.$$

That is, a rise in Mo of 100 causes Y to rise by 66.7, due to the induced rise in consumption spending.

The most serious simplification in our model as presented concerns the foreign sector, which we have treated as fixed and given, with foreign trade balanced (except for one intermediate period), and with a marginal propensity to import of zero. If we relax this by introducing a positive marginal propensity to import, then our analysis is complicated in two ways. In the first place, the money supply is no longer given, but will change each period by an amount equal to E - M. Since imports are now a function of income, it follows that the rate of change of the money supply is also a function of income. The problem of incorporating this fact into our analysis is that we are working jointly with stocks and flows. If the money supply starts out at x at the beginning of the period and rises by dx over the course of the period, it is hard to know what value to use as the determinant of spending during the period. There is no fully accurate approach to this difficulty which is empirically feasible. We shall work primarily with the average of original and final values of Mo, although in some cases we shall also find it necessary to look at the pattern and causes of change within the period.

The incorporation of the positive marginal propensity to import into our income equation is straight-forward. We know that the level of domestic income or product is given by

$$Y = C_h + I_h + E,$$

where C_h and I_h denote expenditure on domestically-produced consumption and investment goods. This relationship could also be expressed as

$$Y = C + I + E - M,$$

where C and I now include total expenditures, whether on imports or on goods made at home. Now if we write

$$M = mpm\ Y + n,$$

where mpm is the marginal propensity to import and n is a constant, then

$$Y = (cY + bMo) + \bar{I} + E - (mpm\ Y + n)$$

$$= (bMo + \bar{I} + E - n)/(1 - c + mpm).$$

The intuitive explanation of this is similar to that given above; the numerator tells the "original" proportionate increase in consumption spending; the denominator, the extent to which this increase will be multiplied by respending. Now, however, we have an additional "leakage," expenditure on imports.

Before going on to further complications and greater realism, let us point out that, in the model as we have presented it, there is always a tendency for foreign trade to move toward balance. If the original equilibrium gives a trade surplus, the money supply will grow each period; the effect of this would be to increase the propensity to spend, presumably on both domestic and imported products. This movement would continue each period until the point was reached where the money supply ceased to grow, i.e. where trade is balanced. Similarly in the case of a deficit, the money supply would fall each period, leading to lower expenditure and lower equilibrium income. Again, a part of the expenditure change would be on imports, leading towards trade balance; this procedure would continue, other things being equal, until exports equal imports.

The next step is to introduce the commercial banks into our model. We know that, when a bank makes a loan, it increases the public's assets and liabilities in equal amounts. However, the fact that the new assets can be spent immediately while the liabilities will not fall due for some time means that the public will feel itself to be in a more liquid situation. If we look only at the public's spendable assets, bank lending increases them, while the repayment of bank loans reduces them. Banks cannot change the <u>net</u> monetary situation of the public (assets minus liabilities), but they can change the <u>gross</u> position, and for many purposes this is more important.

This means that the prevailing monetary situation can no longer be taken as fixed and given. On the one hand it will be influenced by variables determined within the system (i.e. the balance of international trade); on the other, it will be affected by variations in the level of bank lending. It is at this point that our analysis becomes interesting from a policy point of view; for these are variables which might be subject to administrative control. One of the goals of our study is to examine whether such controls would provide a useful and feasible method of influencing the economic system in East Africa.

The next step in generalizing our model concerns the treatment of other monetary assets besides cash and demand deposits. Time and savings deposits are often classified as quasi-money; besides these, we are concerned with liquid monetary assets held overseas. Under the currency board system, these funds can be transferred freely and at will between East Africa

and London; in the East African situation, one might argue that they could best be included in the category of quasi-money.

If we are interested in testing the hypothesis that the public's holdings of monetary assets influence their pattern of expenditure flows, then it seems wise to use a quite wide concept of monetary assets. We propose to work with a total which we shall call gross monetary assets of the public (denoted GMo), consisting of money (currency and demand deposits), quasi-money, and private foreign liquid assets.[1] Changes in the level of this total would be affected by four factors: 1) net[2] new bank lending; 2) net[2] private borrowing and grants received from overseas; 3) the balance of trade surplus; and 4) the operations of the government. There is a fifth factor, price changes, if we are concerned about the real value of the money stock. These five together determine the total gross amount of monetary assets in the hands of the public. It is this total which the public will allocate between holding money, quasi-money, and foreign assets; it is also this total which might affect the public's spending propensities in the way developed in the course of this section.

Following the pattern which we have set out, then, we hypothesize that the level of consumption is influenced by the quantity of gross monetary assets in the hands of the public. We can reformulate this relationship as follows:

$$C = cY + bGMo$$

$$C/Y = c + bGMo/Y .$$

In the following section we would like to examine the data for the East African territories, to see if they give any support to the hypothesis that there is such

---

1. The inclusion of foreign asset holdings in the category GMo is rather debatable; their inclusion could only be fully justified by a thorough examination of the nature of these assets, their ownership, and the purposes for which they are held. We have found very little information in this regard; therefore, although we shall work primarily with the total GMo including these assets, we shall also examine the data on domestic monetary assets alone.

With regard to direct local investments, (where savers lend directly to local borrowers), it might be argued that they should be included too; we exclude them for two reasons: 1) as the claim of the lender on the borrower is often not easily shifted, the net increase in liquidity is probably not great; 2) in East Africa the extent of this direct lending is probably fairly small, and in any case is unmeasured. In the case of foreign assets, if these are real goods (e.g. inventories), they could best be included in the balance of trade figures; if non-liquid investments, as an offset to foreign borrowing (giving a net figure, borrowing minus lending); only if the stock is of liquid monetary assets should we include them in gross monetary assets. In practice, we are not able to separate out private financial investments in London; these are included in GMo, along with liquid sterling assets.

2. Net in the sense of new loans minus repayments, not in the sense of increased assets minus increased liabilities.

a monetary variable in the consumption function. Before proceeding to that examination, however, we should like to say a word about the way in which prices enter into our analysis.

Our discussion so far has been based on an assumption of constant prices; the relationships could be thought of in either real or money terms, since these were equivalent. Now we must take account of the situation where prices do change. Our goal in this regard should be to make sure that we are working with a consumption function which has no money illusion in it.

In a consumption function of the form $C = cY$, there is no problem; consumption is a constant fraction of income, and will stay constant in real terms if Y does but prices change. Once we introduce an intercept into the function, however, the situation is different. If we work with a function of the form $C = cY + k$ and k is constant in money terms while prices change, then its real value will alter, implying a spurious change in the real value of C.

In the case of a consumption function with a monetary variable such as the one with which we have been working, the fall in the real value of the final term (bMo) is not a spurious one, as in the case of the k intercept; rather it is an accurate reflection of the fall in real value of the given money holdings as prices rise. If we use the consumption function in the form $C = cY + bMo$, expressing all terms in current values, even when prices change, this implies no money illusion.

In actual fact prices in East Africa have not fluctuated widely during the period of our study. Available price indices are of quite limited usefulness, since they measure only cost of living in the principal city of each territory; from 1954 to 1959 the increase ranged from 12% in Nairobi to 15% in Kampala. As we indicated above, however, our consumption function is a theoretically reasonable hypothesis even over the course of price fluctuations; for the purposes of the present analysis, we do not need either to correct for price changes or to assume price stability.

4. Analysis of the Data

This section is essentially a commentary on the tables of Appendix A. As the notes to those tables indicate, they have been built up on the basis of a number of specific assumptions, which have been forced on us by the statistical data available, and which should be born in mind in interpreting the results. In this section we shall ask what they indicate with regard to our discussion in section 2, assuming that the tables do present a reasonably accurate picture of the economies of the three territories.

It is unfortunate that we have only been able to construct complete tables for the three-year period 1956–58; this limitation is due to the fact that a large part of our foreign balance data is taken from the official balance of payments estimates, which are only available for those three years. While this shortness of period sharply limits the extent to which our data might "prove" anything, we feel that they can still give a reasonably clear indication of patterns and trends. In many cases we have been able to extend the

period by indicating what such data as we have show for earlier and later years.

Let us recall the hypothesis which we set forth in section 2; it is that there is a positive correlation between the average propensity to consume, C/Y, and the level of monetary assets outstanding, which we have called GMo. We specified this relationship as follows:

$$C = cY + bGMo$$

or

$$C/Y = c + bGMo/Y$$

What we would like to know, then, is whether the value of b is consistent enough to be significant, and large enough to be of interest for policy purposes.

Sophisticated statistical techniques are not likely to be of much help to us here. It would not do to mix the three territories, since they might be quite different in their reactions to monetary changes; this leaves us with only three observations for each territory. With such a small sample, we can best proceed by an intuitive approach. Let us look at the data on a country-by-country basis, to see what they show.

A. Uganda

The outstanding characteristic of this territory, which the data show for our three-year period but which seems to have been true over a much longer time, is the continuing increase in the monetary assets of the residents of the area. Set against this is the surprising constancy of internal monetary assets, once again for a longer span than our three-year period. For example, at the beginning of 1954, domestic monetary assets stood at £26.2 million; the table shows that, by the beginning of 1959, after some fluctuations, they still were only £26.5 million. Over the same period the domestic monetary product rose by nearly 15%. During the three of those years for which we are able to make estimates, the total stock of monetary assets in the hands of the public (including liquid assets held overseas) rose by £21 million; such information as we have indicates that, for the preceding two years, the rise was equally impressive. If our data are accurate, the implication is that the residents of Uganda were accumulating substantial monetary assets outside the country, probably mostly in sterling securities.

Turning to the pattern of consumption spending, it is difficult to find any reflection of these monetary developments there. The year when the propensity to consume was the lowest was precisely the year in which monetary assets increased most sharply. The pattern in the first two years at least moves in the right direction; consumption is higher when monetary assets are higher, both expressed as a ratio to disposable income (see Part III of the tables). However, when this is seen as a part of the continuing upward trend in monetary assets, we cannot put any weight on the fact that, for one period, the consumption ratio also rose.

Our information on changes in external asset holdings is obtained only indirectly, as the difference between implied changes in over-all asset holdings and changes in the level of identifiable (internal) assets; the analysis

itself gives no indication of who holds these assets, in what form, or for what purposes. Once again leaving aside the possibility of major statistical errors, the most likely explanation is that these assets were in the hands of expatriate individuals and firms. In view of the peculiar pattern of change in this rather uncertain element of the data, it seems wise to look also at the relationship between consumption spending and holdings of domestic monetary assets alone. The data for this examination, which are given in Part III, line C of Appendix Table A on Uganda, are more stable than those based on over-all monetary assets, but are no more revealing of any interrelationship between monetary developments and expenditure patterns. Neither of these approaches lends any support to our hypothesis.

It is difficult to assess the importance of the banking system in determining the level of monetary assets in Uganda. We must first recognize that our three-year period, with its very small variations in bank lending, is somewhat unrepresentative in this regard (see Part I, item 2B of the table); bank lending increased by £3.3 million in 1955. However, it seems clear that the foreign trade balance predominated over other determinants of the level of monetary assets. It is therefore somewhat surprising to find that, if we compare bank loans to either demand deposits, total deposits, or total domestic monetary assets in the three East African territories, the ratio is substantially higher for Uganda than for either of the other two territories, as the following table shows:

Ratio of bank loans, advances, bills discounted and local investments to:	Kenya	Uganda	Tanganyika
Total deposits (demand, savings, time)	.69	.76	.51
Money (demand deposits, currency)	.53	.67	.43
Local monetary assets (money + quasi-mon.) (figures for Dec. 31, 1958)	.41	.49	.32

In a sense this is simply a reflection of the fact mentioned above, that there seems to have been a substantial and continuing increase in the external asset holdings of the residents of Uganda. Since we can only indicate changes in these external holdings, and not their total levels, we have no way of comparing the ratio of bank lending to total monetary assets (internal and external) in the three territories; if we could do so, then presumably Uganda would come out lower, and the place of bank lending in her monetary system would be more accurately seen. Furthermore, the ratio of money supply to monetary product is substantially lower in Uganda than in the other two territories, probably due in large part to the lesser importance of expatriates in that territory. Since the organization of the commercial and industrial sectors there is not greatly different from that of Kenya and Tanganyika, the higher ratio of loans to money supply is not unreasonable.[1]

The tentative conclusion to which we are led, then, is that bank lending

1. I am indebted to Mr. R. H. Green for suggesting this interpretation.

seems to be a fairly important source of domestic monetary assets in Uganda, although including external assets as well might temper this position; but we can find little support for the contention that consumption spending is influenced by the level of monetary assets.

B. Tanganyika

On first glance it seems that the data for Tanganyika are no more encouraging. Closer examination indicates, however, that our use of averages of original and final values of GMo conceals monetary developments within the period which might be of some importance in our analysis. Let us examine the figures for Tanganyika in greater detail.

At the beginning of 1956, GMo stood at £32.5 million (assuming original external assets to be 0). During that year, due primarily to the favorable trade balance, this figure rose to £36.0 million. In the following year imports rose substantially while exports fell, so that the over-all private trade balance deteriorated by £12.3 million. Other factors tended to cushion the impact on the monetary system of this shift (e.g. bank lending rose, the government moved into a deficit position), so that the total monetary assets only fell by £3.1 million by the end of the year. In 1958, an improvement in the trade balance was offset by the fall in bank lending and other factors, with the result that the level of monetary assets remained nearly unchanged.

The difficulty with our simplified monetary approach, which looks at the average of original and final values of GMo, is that it treats 1956 (when trade was favorable and monetary assets rising) as being the same as 1957, when the opposite was true. It would seem more plausible to argue that it was the high level of assets outstanding at the end of 1956, caused by the high export earnings during that year, which caused the consumption ratio to rise so sharply in 1957; that a part of this increased consumption spending went for imports, thereby adding to the deterioration in the foreign trade balance in that year, which in turn brought its own correction by causing the monetary ratio (GMo/Y) to fall during the course of 1957.

It should not surprise us that, with such a short period and with our rather crude data, we cannot prove that this is or is not the proper interpretation. What we can say is simply that this interpretation is consistent with the data, and that if it is the proper one, it does imply a positive relationship between monetary asset holdings and the level of consumption spending.

Turning to the question of the role of the banking system in determining the level of monetary assets, let us refer again to the table on page 82 above. We see from that table that bank lending as a percentage of monetary assets was the lowest of the three territories in Tanganyika; that is, bank lending was a less important source of monetary assets there than in the other two territories. Against this must be set the figures in Part I, line 28 of the appendix table, which indicate that, although the total may have been relatively small, fluctuations in this total were quite important in determining the over-all change in the monetary situation in any one year. The rise in bank lending in 1957 helped to offset the deterioration in the foreign trade balance, moderating the

fall in monetary assets during that year; in 1956, the opposite situation prevailed.

As in the case of Uganda, we must recognize that the predominant determinant of the monetary situation was not bank lending but the wide fluctuations in the foreign trade balance. The details of this trade, over a somewhat longer period, are given in the accompanying table. There are two things of interest to note from this table. The first is that fluctuations in the imports of consumer goods formed a substantial part of the fluctuations in total commercial imports. The second is that, with the exception of the import "bulge" in 1955, imports as a whole — and imports of consumption goods as well — seem to have been strongly influenced by the level of export earnings of the preceding year.

In part this can be explained as a lagged income effect, without bringing in any monetary parameters; yet it could also be argued that, as the link which joins the higher income of one period (from higher exports) to the higher consumption spending of the next, the monetary system has a crucial role to play; specifically, if we could moderate the rise in monetary assets in times of export booms, by controlling bank lending, for instance, then perhaps we could also moderate the rise in consumption spending in the following period.

Tanganyika's Visible Foreign Trade

	1954	1955	1956	1957	1958	1959
Total exports	37.8	37.4	46.3	41.0	43.8	47.2
Commercial imports	27.9	38.2	30.8	35.6	30.6	31.2
Consumption imports [a]	13.0 [b]	15.9	11.8	13.6	11.7	13.0
Trade balance	9.9	-0.8	15.5	5.4	13.2	16.0
Export price index	100	87	91	88	81	85
Export quantity index	100	115	136	124	142	147
Import price index	100	96	97	95	95	94

All trade figures in millions of pounds. All figures (except the indices) are for private (i.e. non-governmental) trade only, with all countries outside East Africa.

[a] Includes consumer goods, food, drink and tobacco.

[b] Estimated.

Source: East Africa High Commission, East African Statistical Department, Quarterly Economic and Statistical Bulletin, various issues.

Probably all observers would agree that some kind of lagged income effect accounts for a part of the wide fluctuations in the average propensity to consume during our three-year period; it takes time to decide on, order, and transport imported consumer goods to East Africa. However, between 1956 and 1957, imports of consumer goods rose by only £1.8 million, while total consumption spending increased by £14.3 million; the increase seems to have been primarily in spending on local goods, where the explanation of such a long lag

behind income changes in consumption spending is less plausible.

The conlusion to which we are led—and we must emphasize that it is quite a tentative conclusion—is that in Tanganyika the level of monetary assets seems to have had some effect on the level of consumption expenditure; that while changes in the trade balance are likely to continue to be the most important single determinant of GMo, bank lending can play a role in modifying the extremity of these fluctuations. If a monetary authority could induce or force the banks to vary their lending in such a countercyclical way, it might prove helpful in stabilizing the demand for consumer goods and the foreign exchange balance of the country. We shall have more to say on this point in the following sections, where this conclusion is modified somewhat by an examination of the specific purposes for which banks make loans, in Tanganyika as well as in the other two territories.

C. Kenya

The place of the commercial banks in the monetary system of Kenya was midway between that of Uganda and Tanganyika (see the table on p. 82 above). If we could take account of the substantial (implied) external assets of Uganda, then Kenya would probably emerge with the highest ratio of bank lending to monetary assets, which is what we would expect in view of her more developed economy. Furthermore, annual fluctuations in the total of bank lending have been much larger in Kenya than in the other two territories; in 1954 and 1955 bank lending rose by over £10 million a year, while the largest figure for either of the other territories was £3.6 million (for Tanganyika, in 1959). During our three-year period 1956–1958, the level of bank lending was the most widely fluctuating element among the determinants of GMo in Kenya.

While individual sources of monetary change varied rather widely in this territory over the three-year period, the total level of monetary assets, both internal and external, showed a rather surprising constancy. In view of the crudeness of the data, it would be foolish to rely heavily on changes which only show up in the third decimal place. Such changes as the data do show were in the "right" direction (i.e. consistent with our hypothesis); but with such small changes and such a short period, we must refrain from attempting to use these data to "prove" that the hypothesis is valid. By the same token, they would be equally unable to refute it; the changes were too small to give a good indication either way.

D. Additional Considerations

One might ask whether changes in income distribution had any relevance to our analysis. If the poor have a higher marginal propensity to consume than those who are better off, and if there has been a shift in income in favor of poorer groups in the economy, then we would expect there to be a rise in the average and marginal propensity to consume as this shift of income takes place. In the context of East Africa this approach involves a great deal of conjecture, since we know almost nothing about consumption patterns of different categories of income recipients in that area. In Kenya there was a noticeable increase in

the portion of total money income going for paid employment (i.e. wages, as opposed to other types of factor income), this proportion rising from 53.7% in 1956 to 55.6% in 1957. On the basis of the type of argument set out above, this shift might have caused some increase in the average and marginal propensity to consume of the economy as a whole; one could then argue that this rise was approximately offset by the fall in the ratio of money to income in this period, so that the average propensity to consume remained nearly constant (see Appendix A, Part III, for Kenya). In 1958 the slight decrease in the share of paid employment in total income (to 55.0%) as well as the slight fall in the monetary ratio would then account for the drop in the average propensity to consume. In Tanganyika, the share of value added in "rural" sectors (agriculture, livestock, forest products, hunting and fishing) fell sharply, from 46.4% of the total in 1956 to 39.9% in 1958. If it could be shown that income recipients in these rural sectors have a lower marginal and average propensity to consume (money income) than the rest of the economy, then this shift might account for the implied upward trend in our figures for C/Y during this period.

It should be clear that this discussion is highly conjectural; we have no real knowledge of expenditure patterns of separate groups in East Africa. The explanations given seem reasonable, and are consistent with the data; only further empirical work will show whether in fact they are valid.

In principle one could make a similar argument concerning the results of changes in real income per capita; however, the price indices by which one would want to deflate gross product to express it in real terms are so weak that we do not feel that anything significant can be added through this approach. As an indication of the difficulties which one would encounter along this line, we refer to a study of the Uganda Unit of the East African Statistical Department; after taking account of the fall in export prices and the rise in domestic wage rates, they estimate that

> the Gross Domestic Product in 1959 at 1954 prices might therefore be of the order of 5% higher than the value given at 1959 prices . . . The calculation of Gross Domestic Product at constant prices is designed to measure the country's economic effort. It in no way implies that the purchasing power of the income resulting from the effort has changed in the same proportion as output at constant prices since changes in the terms of trade have to be allowed for . . . In addition there have been considerable shifts in purchasing power within the economy, employees becoming better off while peasants face lower incomes.[1]

There is the further problem of knowing what population figure we should use in computing income per capita. Since we are working only with money income, we should exclude that portion of the population in the "subsistence sector."

---

1. East Africa High Commission, East African Statistical Department, The Gross Domestic Product of Uganda: 1954-1959 (Entebbe, Uganda: Government Printer, 1961), p. 34.

However, as we indicated earlier, the dividing line between the two groups is not at all precise.

With all these difficulties facing us, it does not seem that we can say anything precise enough about changes in real income per capita to add to our understanding of the pattern of consumption spending; as an approximation, we feel that it is reasonable to neglect changes in this variable either as a factor differentiating one territory from another, or as a factor of significant change during the period within any one territory.

Of the three territories we have examined, the data in one (Tanganyika) are consistent with the hypothesis that the monetary situation had an influence on the level of consumption spending; in a second (Uganda) the data are quite inconsistent with this hypothesis, while in the third (Kenya) the changes were in the "right" direction, but they were so small that they give little support to the hypothesis. If we take account of shifts in income distribution on the basis of seemingly plausible but unverified assumptions, then the agreement between the data and our conjecture is better in both Kenya and Tanganyika. In concluding this section, perhaps it would be worthwhile to point out two possible reasons why the data do not give more support to the hypothesis. The first concerns the importance of expatriates and their financial dealings in the economic life of East Africa. In our empirical work we have not been able to separate liquid assets held in London from investments overseas; both are in fact included in our category GMo. If a British businessman in East Africa finds his stock of money to be "too high," his first reaction would probably be to increase his investments in London. Only to the extent that he considers his holdings of money-plus-London-investments to be too high would he vary his consumption spending, or his investments in East Africa, in a way which would show up in our analysis.

It would have helped in this regard if we could have separated the income, expenditures, and money holdings of permanent residents of East Africa (of whatever race) from those living there on a temporary basis. This information is not presently available, and is not likely to be obtainable in the near future. The fact that we have had to work with all groups in the country together, and that we have not been able to refine our category of monetary asset holdings as sharply as we would have liked, is probably a major weakness of our analysis.

A second and closely related limitation is that we have not been able to separate money holdings of individuals from those of firms. To the extent that individuals are shielded from monetary fluctuations by changes in money holdings of firms, we would expect the monetary impact on individuals' consumption patterns to be reduced. If this is the case, then we should go on to examine the relationship between business spending and monetary developments. This is the subject matter of the following section.

5. Enterprises

In the preceding section we have discussed at some length the possible impact of monetary parameters on the level of consumption spending. We

would now like to proceed to an examination of the relationship between monetary variables and the activities of producers. There might be such a relationship at either or both of two levels: at an aggregative level, as reflected in the total of GMo which we have worked with in earlier chapters; or at a more detailed level, particularly with regard to the level of bank lending.

Starting with the more aggregative approach, there are three reasons why one might look for a relationship between the ratio of monetary asset holdings to income on the one hand and the level of business activity on the other. In the first place, as we indicated at the conclusion of the last section, we have not been able to separate money holdings of firms from those of individuals; we are working with a composite of these two. To the extent that the rise in the ratio of money to income is reflected in the money holdings of firms, these firms might be induced or allowed to expand their production levels by this increase in their liquidity. If the rise is in the money holdings of individuals who react by seeking to buy new securities, this also might allow firms to expand. Finally, if individuals react to a rise by increasing their consumption spending, this in turn might lead producers to increase their output. The first two reasons imply that shortage of finance is a limiting factor in the expansion of production; the third implies that it is inadequate demand for the product which restricts output.

If for these reasons or others there is a relationship between GMo/Y and business activity, we might ask what aspects of business activity would respond. One would want to look first at total production, perhaps leaving aside the agricultural and government sectors, since these are sectors in which production is largely determined by other factors, at least in the short run. Since the over-all supply elasticity in the remainder of the economy is likely to be rather low in underdeveloped countries, there might be a considerable lag between the stimulus to change and the adjustment in output. If this were so, we could get an indication of the extent to which firms are seeking to change their production levels by looking at the current levels of capital formation, as well as at two of its principal components, imports of capital goods and output in the construction industry.

While this line of reasoning seems reasonably straight-forward, the data with which one would want to verify it are not particularly revealing.[1] All the problems which we met before of a too-short time period, of knowing what time dimension to select, of overaggregation of data, return to the fore. As they stand, the data indicate no clear relationship between monetary parameters and the level of business activity. Let us go on, then, to a closer examination of more disaggregated data.

Our disaggregation will be along two main lines. On the other hand, we would like to concentrate on one of the determinants of GMo which is of special interest to us, namely, the level of bank lending; on the other, we would like to look more closely at the composition of the expenditure and production flows which are financed by this bank lending.

---

1. See Appendix D of the dissertation on which this essay is based.

The relative importance of the various categories of bank lending in each territory is shown in the following table, as of the end of 1958:

	Kenya		Uganda		Tanganyika	
	£ mil.	%	£ mil.	%	£ mil.	%
Loans and advances	29.1	81	11.3	88	9.1	87
Bills discounter	5.1	14	0.6	5	1.3	12
Investments in E. A.	1.8	5	1.0	8	0.1	1
Total	36.0	100	12.9	100	10.5	100

Source: Quarterly Economic and Statistical Bulletin

It is clear that the main component of bank credit in East Africa is loans and advances. The prevalence of Indian traders means that the practice of discounting bills is more widespread there, and particularly in Kenya, then in most underdeveloped areas; still bills discounted comprise a relatively minor portion of total bank credit in all three territories. The other main type of bank credit is through local investments. This item consists primarily of bank holdings of East African government bonds; as the table shows, it is of rather limited importance.

In East Africa, as in all areas where British traditions predominate, the main function of bank lending is considered to be the provision of short-term finance, preferably of a "self-liquidating" nature. Most bankers fully realize the limited meaning which can be attached to this term; in the course of a business downturn, even the most liquid of physical assets might be realizable only at a loss. However, the principle remains, and banks still seek to channel their credit primarily into directions where it will be used to finance the completion of an economic transaction. Although banks are now more ready than formerly to finance a limited amount of medium-term credit, it still seems fair to say that the great majority of loans and advances in East Africa are for such short-term financing purposes.

The distribution of loans and advances among various sectors of the economy was as follows, as of the end of 1958:

	Kenya		Uganda		Tanganyika	
	£ mil.	% of tot.	£ mil.	% of tot.	£ mil.	% of tot.
Industry	4.6	16	4.7	42	2.0	22
Agriculture	6.2	21	1.3	12	2.6	29
Other	18.3	63	5.3	47	4.5	49
Total	29.1	100	11.3	100	9.1	100

Source: Quarterly Economic and Statistical Bulletin

In all three territories the largest category of loans was that entitled "other"; this consists primarily of loans to finance internal trade and

imports.[1] It should be pointed out that virtually all of Uganda's imports come in by ship to Mombasa and are transported by rail or truck across Kenya; a significant portion of loans to importers to help finance this trade is granted by banks in Nairobi and Mombasa, with the result that Kenya's figures are increased while those for Uganda are reduced.

Another measure of the relative importance of bank lending to the various sectors is to indicate the level of bank lending for a given type of activity as a proportion of the value of transactions in that sector. In the case of agriculture and industrial loans, we can use value added in the corresponding productive sectors as a measure of the value of these transactions. Gross value of product (including intermediary inputs), which might give a better measure for this purpose, is not available. With regard to the third category of advances, if the explanation as to their nature given above is correct, then the best measure of the transactions which the advances help to finance might be the sum of commercial imports plus value added in the domestic wholesale and retail trade sector. On this basis, we can draw up the following table for the year 1958.

Loans as a Percentage of Transactions by Sectors,* 1958

	Kenya	Uganda	Tanganyika
Industry	14	19	15
Agriculture	18	4	6
Other	24	12	9

*Loan figures are average of the four quarterly figures for 1958.

We have already commented on the figures for "other loans"; the fact that Kenya is the financial center of the area accounts for the higher figure for that territory. The figures for agriculture are quite revealing; they are a clear reflection of the fact that in Uganda, and to a lesser extent in Tanganyika, agricultural products are grown primarily on small-scale, African-owned farms; in Kenya, on the other hand, the majority of the agricultural cash products are grown on large-scale farms owned and operated by foreigners, who are used to turning to the banks for finance, and who are considered by the banks to be more credit-worthy. The pattern of lending for industrial purposes is not so easy to explain. Commenting on the figures for 1950, Newlyn and Rowan attributed the relatively high level of industrial lending in Uganda to "the greater reliance on bank credit relative to working capital resulting from the small scale of enterprise."[2] This explanation does not sound convincing, particularly in view of the fact that quite a substantial

1. W. T. Newlyn and D. C. Rowan, op. cit., p. 86. Their study has been quite helpful on several points discussed in this chapter.

2. Ibid., p. 88.

portion of industrial production in Uganda is accounted for by a few large establishments, often with government participation through the Uganda Development Corporation. It would seem more plausible to argue that it was these relatively few large enterprises which received most of the loans for industrial purposes in Uganda. These loans were substantially less important compared to total bank lending in the other two territories, although relative to output of the industrial sector they were not greatly different from Uganda.

When we examine the relationship between loans and transactions on the basis of this sectoral breakdown over a period of several years, we find relatively little stability in this relationship (see Appendix B). For example, there were quite substantial fluctuations in the ratio of industrial loans to value added in the industrial sector in Tanganyika; and similarly for other territories and other sectors. The detailed figures underlying these ratios show that the fluctuations are due to changes in both lending and production levels.

One must be rather careful about conclusions to be drawn from this examination. It certainly does not follow that bank lending and production totals are completely unrelated, in the sense that if banks were forced to restrict their lending by 25% then production totals would be unaffected. The data do imply that there are other sources of finance available to producers and traders, and that the relative importance of these other sources has varied fairly substantially from year to year. In this situation the difficulty of monetary policy—of attempting to influence trade and production totals by controlling the level of bank lending—is not that it is impotent, but rather that it is unpredictable. It still seems reasonable to expect that, by forcing or inducing banks to restrict their lending, we could hold trade and production to lower levels than they would otherwise have reached. The problem is that it is quite difficult, with the data at hand, to know how powerful a tool we are using, how strong a reaction we should expect from a given change in the level of bank lending.

To some extent this same problem faces the monetary authorities in the United States as well. This difficulty is partially offset in this country, however, by the fact that one of our major instruments of control, open market operations, can be varied continuously and in small steps, without discontinuities or announcement effects; we can gradually tighten the market until a certain desired degree of restrictiveness is attained, see what the reactions are in the production sphere, and adjust accordingly. In East Africa such a fine instrument of control is not available, and is not likely to be developed in the near future. In this situation the difficulty of knowing in advance the result of any monetary policy decision is likely to be a more serious limitation on the effective use of monetary policy than in our country.

In this chapter we have been concerned with the sectoral breakdown of bank lending in East Africa, as well as the relationship between changes in the level of lending and corresponding production and expenditure flows. We must now attempt to relate these findings to the discussion in earlier sections, taking an over-all look at the feasibility of using monetary policy as an instrument of control in East Africa.

6. Conclusions

A. General Considerations

Before presenting a summary of our findings, we should like to say a word about the over-all goals which a monetary authority might set for itself. In developed economies, these goals might be stated in general terms as follows: 1) to maintain a level of aggregate demand for local products which would just keep the resources of the economy—the factors of production—fully employed; 2) to maintain some balance among the various uses of the product (consumption, investment by sectors, and government) consistent with long-run goals of rates of growth and size of the government sector. We believe that, in these general terms, the goals for underdeveloped countries are the same; but for these countries there arise problems of interpretation and implementation far greater than for developed countries. It is far more difficult, for example, to give any meaning to the term full employment. Furthermore, the largest demand fluctuations are likely to be in a sector which is beyond government control, namely, the export sector. A government authority desiring to keep the total demand for local products constant would have to try to induce countercyclical variations in expenditures in another sector. In view of the commonly asserted immobility of factors of production in underdeveloped countries, this path seems rather fruitless. Perhaps the most that a government authority could hope to do in such a situation would be to try to keep the domestic demand for home-produced goods at the maximum level which the economy could meet satisfactorily, in view of its susceptibility to demand-induced inflation and the dangers of a foreign exchange shortage.

Monetary policy is only one of a number of tools with which the government authorities might pursue this goal. Since the level of disposable income is one of the main determinants of aggregate domestic demand, fiscal policy could have an important role in tempering fluctuations, as could an active stabilization policy of the marketing boards.[1] From the point of view of income stabilization, monetary policy could probably best be looked upon as a supplement to these other stabilization tools.

On the further point of the allocation of output among various sectors in accordance with long-run goals, the possibility of influencing the allocation in this way depends primarily on the number of tools of control available to the authorities. If they have only one "handle to turn," and can only increase or decrease demand by this one method, then they have no freedom to discriminate among sectors; but if there are several instruments of control for

---

1. Kennedy, Ord and Walker have argued that these two alone have been quite successful in stabilizing income in Uganda; in Kenya and Tanganyika, on the other hand, the marketing boards have been far less important in their income stabilization roles, and export taxes are negligible. See "On the Calculation and Interpretation of National Accounting Material in East Africa," Paper read to the meeting of the International Association for Research in Income and Wealth, Addis Ababa, January 1960 (mimeographed).

them to use, then they might bring the aggregate demand in the economy to a desired level in any of several ways, by encouraging one sector and discouraging another, or vice versa. The extent to which they can do this depends on the variety of choice among different methods of control, influencing different sectors of the economy. This takes us into questions of the efficacy of fiscal policy, tariff policy, and many other such tools of government control. In this whole complex of issues, we shall restrict ourselves to the following question: if monetary policy is a feasible tool for influencing the level of expenditure and production in the economy, on what sector or sectors does it impinge? Taken along with an appraisal of other instruments of control, this discussion should permit the authorities to know the extent to which they can affect not only the total of aggregate demand but its allocation among sectors.

B. The Possibilities of Controlling the Monetary System

Let us summarize our findings of section two above, where we discussed various methods by which a government authority might hope to influence the monetary situation in the economy. The main points to be mentioned are as follows:

1) Of the standard tools of monetary control, open market operations and a restrictive discount rate policy are likely to have only limited effectiveness in East Africa due to the international character of the banks and the limited capital market. An expansionary discount rate policy might be more successful in helping the banks to expand their lending in East Africa.

2) Direct controls of bank lending, which set a maximum limit each period beyond which banks are not permitted to go, seem possible but of limited usefulness, and difficult to apply administratively.

3) With regard to standard reserve requirements, the international character of the banks means that changes in reserve requirements would have to be more substantial before any induced changes in the level of lending would occur. A possible variation of this approach would be to make the required deposits a function of the banks' domestic lending. Instead of working through the aggregate balance sheet, these deposits would affect the profitability of local lending directly, and hence might be more likely to influence the banks' willingness to lend. If the goal were to bring about an expansion of loans, interest might be paid on these deposits. Furthermore it should be possible, in setting either deposit requirements or interest rate payments, to discriminate between different categories of loans, thereby permitting the authorities to be selective in their restrictive or expansionary pressures.

4) Other financial institutions besides commercial banks are likely to be of growing importance in the financial system of East Africa. Because several of these institutions are under direct government supervision at present, it should be relatively easy to extend this control to include a goal of monetary stabilization. Since a great many of their loans are for purposes directly related to the development process, however, it would seem wiser to encourage their lending at all times, pursuing stabilization goals through other channels.

C. The Impact of Monetary Variables on Expenditures

Here we need to examine each territory separately, since the conclusions differ somewhat.

1) Uganda. In this territory the figures imply a regular and substantial increase in the external asset holdings of individuals, made possible primarily by the continuing export surplus of the territory. Variations in the level of bank lending were an insignificant factor in the determination of the over-all monetary situation. There is no evidence that these monetary developments had any impact on the level of consumption spending.

Looking in more detail at the sectors in the economy which received bank loans, we find that quite a substantial proportion of these loans were for industrial purposes. The category of "other loans and advances," primarily for the financing of imports and internal trade, was also substantial, while agricultural lending, the discounting of bills and investment within the territory were negligible. Fiscal policy and marketing board operations have been fairly successful in bringing a measure of stability to income and expenditures in the territory.

In view of these facts, and in view of the desire of the government to encourage the expansion of the industrial sector, the most reasonable goal of a monetary authority might be to encourage the granting of industrial loans to such an extent that no feasible industrial project is held back for lack of financing. In this situation, the privilege of discounting industrial loans at favorable rates with the monetary authority might prove to be particularly important. If the techniques of control are sufficiently precise to permit the authorities to discriminate among different sectors, they might supplement the stabilization tools mentioned above by inducing banks to restrict their loans for trade and import purposes ("other loans and advances") in times of excessive demand. These are perhaps unexciting conclusions; yet they seem to indicate the most that could be hoped for from monetary policy in the territory in the present circumstances.

2) Kenya. In this territory total monetary asset holdings rose slightly over the period; internal assets (money plus quasi-money) fell sharply during 1956, then remained nearly constant. The level of bank lending fluctuated fairly widely, although it was overshadowed by the foreign sector as the primary determinant of the monetary situation. Unless one relies on an unverified hypothesis concerning the role of income redistribution in the economy, one sees very little relationship between these monetary developments and the level of consumption spending; changes were too small to be able to draw any clear conclusions one way or another.

With regard to the nature of bank lending in the territory, over half of the total was for the finance of imports and internal trade ("other loans and advances"); the remainder was distributed among bills discounted (which probably could best be grouped with "other loans," from the point of view of the uses to which the loans are put), agricultural loans, and industrial lending. Local investments were negligible. Agricultural lending was far more important in Kenya than in the other territories, both in absolute terms and relative to the

level of agricultural production. Since this type of lending goes primarily for financing agricultural improvements and intermediate inputs, any attempt to restrict this type of lending is likely to be self-defeating; the result would probably be a fall in the level of agricultural production. On the contrary, the wisest policy would seem to be the active encouragement of this type of lending, both by the commercial banks and by the Kenya Land Bank.

In view of the predominance of "other loans" and bills discounted in total bank lending, it seems clear that any change in the banks' over-all willingness to lend would be reflected primarily in the extent to which banks help to finance imports and domestic wholesale and retail trade. If a monetary authority forced the banks to restrict this type of lending during a period of excess demand, two results might follow. On the one hand, traders would find it more difficult to bid against one another for the limited supply of domestic goods; on the other, importers would find it more difficult to finance their trade, with the result that the level of imports might fall. Our data are not sufficiently precise for us to be able to judge which of these two effects will be more important. If the shortage of foreign exchange is felt to be less of a problem than increases in internal prices, and if changes in the level of "other loans" affect primarily the level of imports, then a monetary authority might even want to encourage these loans in times of excess demand, to induce a rise in imports. The decision as to which approach is the wisest in this situation would depend on how the authorities view the twin dangers of rising imports and rising internal prices, as well as a more detailed examination of the impact in these two directions of changes in this type of bank lending.

In sum, the authorities would probably want to encourage agricultural lending (and probably the smaller category of industrial loans as well) at all times; in periods of excess demand they might seek to influence the level of "other loans" and bills discounted in an attempt to temper the level of demand and channel it into the least harmful direction. This seems to be the limit of what could be achieved through monetary policy in Kenya at present.

3) Tanganyika. Here internal monetary assets were quite stable. Taking account of implied external asset holdings, total GMo fluctuated somewhat, primarily in response to the trade balance, although bank lending also played a role in this. In Tanganyika there seems to have been a positive relationship between the monetary situation and the level of consumption spending. Our data are not adequate to prove that this relationship was a continuing and important one; however, over the short period which we were able to examine fully, the data do lend support to the hypothesis.

With regard to the nature of bank lending in the territory, once again somewhat more than half of the total was for the finance of imports and internal trade (if we include bills discounted with "other loans and advances"). The remainder was for agricultural and industrial loans. The larger of these two, agricultural lending, was far less important relative to the output of the agricultural sector than in Kenya.

In this situation the authorities might face the same choice which we

discussed in the case of Kenya between permitting imports to rise, thereby drawing on foreign assets, or attempting to hold down the level of imports, with the result that the rise in internal prices might be sharper. In Tanganyika, however, we have reason to believe that the level of bank lending would also have some influence on the demand for consumer goods in the economy. If this is so, then by restricting the level of bank lending in times of excess demand, the monetary authorities might succeed in tempering the over-all demand for consumer goods, as well as reducing the level of imports.

In the East African context, an increase in commercial bank lending can be an effective channel for the inflow of much-needed capital from abroad. In this situation, and excepting the case of excess demand discussed above, it seems that the wisest course for a monetary authority would be to seek to encourage the expansion of bank lending in every way possible. Of course a monetary authority cannot create a vigorous agricultural or industrial sector, and thereby expand the demand for loans from the commercial banks; but by providing generous rediscounting facilities and attempting to develop the local money and capital markets, the authorities can make it easier for the commercial banks to accommodate the demanders who do come to them for loans. This seems like the most important contribution which a monetary authority could make under present circumstances, in Tanganyika as well as in the other two East African territories.

D. Concluding Comments

To a large extent, the results of our examination of the interaction between monetary variables and expenditure flows could be thought of as negative; they imply that certain types of interrelationships which we have looked at are not of great importance in the East African context. We feel that this conclusion can be a significant and useful finding in itself. However, in view of the limitations of the statistics with which we have worked, centering primarily around the overaggregation of our basic data, it might be fair to say that our contribution on this question lies primarily in the method of approach suggested, rather than in the specific results obtained. With the further disaggregation and improvement in accuracy of the basic data, the application and development of our general line of approach should permit a more detailed analysis of the role of monetary variables in the economies of the three territories. These comments would be less true of our discussion of the techniques of monetary control, where we feel that our study is more directly relevent and applicable to questions currently facing the authorities in East Africa.

Events have been moving at an amazing pace in the three territories in recent months. Tanganyika has recently received its independence; Uganda followed in 1962, with Kenya most likely not far behind. On the financial scene, there has been a steady growth in the number of new banks and financial institutions. The head office of the currency board has recently been moved from London to Nairobi, and the board has been completely reconstituted,

with local finance ministers playing a prominent role. It is slowly increasing its fiduciary issue, although in the latest annual report, as of June 30, 1960, sterling assets were still 101% of currency in the hands of the non-bank public.

In an important lecture before the Economics Club of Kenya,[1] Mr. J. B. Loynes, newly appointed member of the currency board and advisor to the Bank of England, pointed the way to the likely course of future change; he indicated that the transition from currency board to central bank could be reasonably smooth, with the currency board itself undertaking an increasing number of the functions associated with central banking. The present allowance of a limited fiduciary issue is a first step in this direction. The next step, still to be taken, involves the use of this provision of fiduciary issue in a flexible, countercyclical way; the currency board would vary its holdings of local government debt instruments to suit the current economic situation. Perhaps it would be accurate to say that, in that stage, the currency board would be underwriting a countercyclical fiscal policy. Our analysis, however, is concerned primarily with a further stage of development, when the monetary authorities begin to try to influence the financial institutions directly. This last type of control cannot be undertaken by the currency board unless it receives additional authority which would make it a central bank in fact, if not in name. While Mr. Loynes' suggestion of a gradual transition seems to be a wise one, we feel that it is also important to look beyond this to the time when monetary control becomes important in its own right. It is at this point that we hope our study will be relevant and helpful.

Appendix A: Data for Monetary and Income Analysis

Uganda

All figures in £ millions

	1956	1957	1958	1959
Part I: Monetary Data				
1. Domestic monetary assets at beginning of period				
A. Money	22.8	21.1	18.3	19.4
i. Currency	7.8	8.2	7.4	7.2
ii. Demand deposits	15.0	12.9	10.9	12.2
B. Quasi-money	4.8	5.7	6.3	7.1
i. Time deposits	1.1	1.7	1.8	1.8
ii. Savings deposits	1.2	1.6	2.2	3.2
iii. Post office savings	2.5	2.4	2.3	2.1
C. Total	27.6	26.8	24.6	26.5

1. J. B. Loynes, "From Currency Board to Central Bank," East African Economics Review, Vol. 7, No. 2 (December 1960), pp. 57-67.

Uganda (cont'd)

All figures in £ millions

	1956	1957	1958	1959
Part I: Monetary Data (cont'd)				
2. Identified cause of change during period				
A. Government deficit	0.6	0.8	0.7	
B. Bank lending in territory	- 0.7	0.4	0.5	
C. Private trade balance (net)				
i. Visible, outside E.A.	15.1	19.8	21.2	
ii. " , interterritorial	- 0.7	- 1.2	- 1.1	
iii. Invisible	- 11.6	- 12.7	- 11.6	
D. Credit and Savings Bk. lend.	—	—	0.1	
E. Marketing boards	0.3	- 0.4	1.5	
F. Total	3.0	6.7	11.3	
3. Total GMo, beginning of period	27.6	30.6	37.3	48.6
4. Average GMo over year	29.1	33.9	42.9	
Part II: Income Data				
1. Social accounts				
A. Sources of goods available				
i. Gross dom. prod. at factor cost (mon.)	103.2	109.8	106.7	
ii. Expend. taxes	6.7	7.1	7.7	
iii. Visible imports	33.2	35.4	34.3	
iv. Invisible imports	13.1	15.0	13.3	
v. Gross dom. prod. at market plus imports	156.2	167.3	162.0	
B. Allocation of goods				
i. Exports	45.9	52.1	52.6	
ii. Gov. spending	22.5	23.9	24.7	
iii. Private capital form.	13.3	11.5	11.0	
iv. Residual: consumption	74.5	79.8	73.7	
2. Determination of income				
A. Gross domestic product at factor	103.2	109.8	106.7	
B. Less depreciation	5.2	5.5	5.6	
C. Net domestic income	98.0	104.3	101.1	
D. Less: Direct personal and corporate taxes	9.6	10.4	8.1	
E. Disposable domestic income	88.4	93.9	93.0	

Uganda (cont'd)

All figures in £ millions

	1956	1957	1958	1959
Part III: Joining				
A. c/Y_d	.84	.85	.79	
B. GMo/Y_d	.33	.36	.46	
C. Domestic monetary assets/Y_d	.31	.27	.28	

Tanganyika

Part I: Monetary Data				
1. Domestic monetary assets at beginning of period				
A. Money	25.7	24.2	23.5	24.5
i. Currency	8.7	9.5	9.6	9.9
ii. Demand deposits	17.0	14.7	13.9	14.6
B. Quasi-money	6.8	8.2	8.1	8.1
i. Time deposits	2.7	3.8	3.5	3.4
ii. Savings deposits	1.5	1.9	2.2	2.5
iii. Post office savings	2.6	2.5	2.4	2.2
C. Total	32.5	32.4	31.6	32.6
2. Identified causes of change during period				
A. Government deficit	- 1.0	0.1	0.2	
B. Bank lending in territory	- 1.4	2.9	- 1.8	
C. Private trade balance (net)				
i. Visible, outside E.A.	15.5	5.4	13.2	
ii. " , interterritorial	- 4.3	- 5.7	- 6.4	
iii. Invisible	- 5.0	- 5.8	- 6.6	
D. Land Bank lending	—	0.1	0.2	
E. Marketing board	- 0.3	- 0.1	0.9	
F. Total	3.5	- 3.1	- 0.3	
3. Total GMo, beginning of period	32.5	36.0	32.9	32.6
4. Average GMo over year	34.2	34.4	32.7	
Part II: Income Data				
1. Social accounts				
A. Source of goods available				
i. Gross dom. prod. at factor	89.3	92.9	97.9	
ii. Taxes on expenditure	7.8	8.8	9.2	
Less subsidies	- 0.2	- 0.2	- 0.2	

Tanganyika (cont'd)

All figures in £ millions

	1956	1957	1958	1959
Part II: Income Data				
A. Sources of goods available (cont'd)				
iii. Visible imports	42.2	47.0	42.6	
iv. Net invisible imports	6.1	7.8	8.1	
v. Gross dom. prod. at market plus imports	145.2	156.3	157.6	
B. Allocation of goods				
i. Exports	48.4	43.1	46.4	
ii. a. gen. government: consumption	14.9	16.3	17.6	
b. gen. government: gross cap. form.	6.4	7.0	7.0	
c. gov. enterprises: gross cap. form.	3.0	2.8	1.5	
iii. Private capital form.	14.0	14.3	14.2	
iv. Residual: Private consumption	58.5	72.8	70.9	
2. Determination of income				
A. Gross dom. prod. at factor cost	89.3	92.9	97.9	
B. Less depreciation	7.1	7.5	7.2	
C. Net domestic income	82.2	85.4	90.7	
D. Less direct personal and corporate taxes	6.3	5.9	4.8	
plus gov. transfer payments	1.8	2.2	2.6	
E. Disposable domestic income	77.7	81.7	88.5	
Part III: Joining				
A. C/Y_d	.75	.89	.80	
B. GMo/Y_d	.44	.42	.37	
C. GMo (original value)/Y_d	.42	.44	.37	
D. Domestic monetary assets/Y_d	.42	.39	.36	

Kenya

Part I: Monetary Data				
1. Domestic monetary assets at beginning of period				
A. Money	78.9	71.0	71.8	67.8

Kenya (cont'd)

All figures in £ millions

	1956	1957	1958	1959
Part I: Monetary Data				
A. Money (cont'd)				
i. Currency	27.0	27.6	28.9	27.9
ii. Demand deposits	51.9	43.4	42.9	39.9
B. Quasi-money	16.3	17.8	19.1	21.0
i. Time deposits	2.8	4.1	5.0	6.1
ii. Savings deposits	3.2	3.9	5.1	6.2
iii. Post office savings	10.3	9.8	9.0	8.7
C. Total	95.2	88.8	90.9	88.8
2. Identified cause of change during period				
A. Government deficit	2.2	1.3	1.1	
B. Bank lending in territory	- 5.1	3.1	- 4.9	
C. Private trade balance				
i. Visible, outside E.A.	- 28.7	- 33.6	- 23.1	
ii. " , interterritorial	4.9	6.9	7.5	
iii. Invisible	6.9	7.5	5.1	
D. Land Bank lending	0.4	0.6	0.1	
E. Private capital inflow	21.5	17.9	10.6	
F. Total	2.1	3.7	- 3.6	
3. Total GMo, beginning of period	95.2	97.3	101.0	97.4
4. Average GMo over year	96.2	99.1	99.2	
Part II: Income Data				
1. Social accounts				
A. Sources of goods available				
i. Gross dom. prod. at factor	146.8	153.4	154.8	
ii. Taxes on expenditure	12.9	12.3	13.9	
iii. Visible imports	73.9	76.5	66.3	
iv. Net invisible imports	- 4.1	- 3.0	- 1.7	
v. Gross dom. prod. at market plus imports	229.5	239.2	233.3	
B. Allocation of goods				
i. Exports	42.0	42.7	46.1	
ii. Gov. spending on goods and services	44.1	46.6	42.2	
iii. Private capital form.	31.1	30.6	27.0	
iv. Residual: private consumption	112.3	119.3	118.0	

Kenya (cont'd)

All figures in £ millions

	1956	1957	1958	1959
Part II: Income Data (cont'd)				
2. Determination of income				
A. Gross dom. prod. at factor	146.8	153.4	154.8	
B. Less: cap. consumption allowance	14.4	15.1	14.7	
Current surplus of government enterprises	3.4	3.7	4.3	
C. Net domestic income	129.0	134.6	135.8	
D. Less: operating and rental surplus of incorporated companies	15.8	15.7	16.4	
E. Plus: Gov. transfers: grants to H.H.	2.9	3.3	3.9	
interest to H.H.	0.1	0.1	0.2	
Business interest and dividends	4.5	4.6	4.6	
F. Personal domestic income	120.7	126.9	128.1	
G. Less: indiv. income tax payments	10.0	9.1	9.1	
H. Disposable personal income (domestic)	110.7	117.8	119.0	
Part III: Joining				
A. C/Y_d	1.015	1.010	.99	
B. GMo/Y_d	.87	.84	.83	
C. Domestic monetary assets/Y_d	.83	.76	.75	

For detailed notes and sources the reader is referred to the dissertation on which this essay is based. We have relied heavily on publications of the East African Statistical Department, supplemented in some cases by correspondence with the East African government officials concerned. It is necessary to make a brief comment on several specific items.

1. Official publications give currency outstanding only as a total for the three territories; we have allocated this total among them in proportion to demand deposits.

2. Government deficit is an estimate of the divergence between total funds raised locally by the government (either from taxes or from borrowing outside the banks) and local expenditures (i.e. excluding overseas debt payments, passage pay, and government imports).

3. Invisible trade is estimated on the basis of the East African Statistical

Department, The Balance of Payments of East Africa, 1956, 1957, and 1958, and T. A. Kennedy, "An Estimate of Uganda Balance of Payments, 1949-1957," East African Economics Review, Vol. 6, No. 1 (July 1959), pp. 1-13. Details are given in Appendix C of the dissertation.

4. In the absence of a detailed breakdown, the entire capital inflow to the area is credited to Kenya.

5. Figures on GMo are based on the assumption that the original external asset holdings are 0.

6. Consumption is a residual; its main weakness comes from the fact that the capital formation figures exclude stock changes.

7. Disposable income is a goal which we have been able to approach with varying success. In Uganda it is simply gross domestic product at factor cost less taxes and estimated depreciation. The figures for Kenya are quite complete in this regard, while Tanganyika falls between the other two territories.

Appendix B: Ratio of Bank Lending to Transactions, by Sectors

	1954	1955	1956	1957	1958	1959	1960
Kenya							
Agriculture	.17	.18	.16	.17	.18	.18	.16
Industry	.17	.14	.13	.12	.14	.14	.15
"Other"	.20	.27	.24	.23	.24	.26	.28
Uganda							
Agriculture	.03	.04	.04	.04	.04	.02	
Industry	.21	.22	.20	.23	.19	.25	
"Other"	.07	.10	.10	.12	.12	.13	
Tanganyika							
Agriculture	.08	.07	.05	.05	.06	.07	.09
Industry	.14	.11	.16	.20	.15	.13	.12
"Other"	.09	.12	.11	.09	.09	.15	.17

Source: East African Statistical Department, Quarterly Economic and Statistical Bulletin.

Summary: The theory of forward exchanges derives its importance from the fact that forward exchange policy may be a useful tool in the fight against two of the most annoying problems of international finance, recurrent speculative runs on currencies and large-scale interest arbitrage of short-term funds. The first part of this essay examines the nature of these two problems and pinpoints to what extent and when forward exchange policy can be used to combat them.

The theoretical model of forward exchange markets developed in the second part of the essay specifies the determinants of demand and supply of forward exchange and the forward exchange rate. For this purpose the behavior of the participants in the forward exchange markets, interest arbitrageurs, triangular arbitrageurs, speculators and traders are analyzed. The resultant model leads to the central hypothesis that during normal times forward rates should be such that interest arbitrage is not profitable and that during times of speculation sizable profit opportunities from artibrage should exist.

This hypothesis is tested by means of computations involving exchange and interest rates from New York, London, Ottawa, Paris and Frankfurt. The hypothesis is found to be essentially correct. Different empirical tests suggest that triangular arbitrage in forward exchange is an increasingly potent force in the markets.

In part four of the essay the model was used to analyze the rationale, costs and hazards of forward exchange policy. Special attention was paid to the influence of triangular arbitrage in forward exchange on the usefulness of the policy. It was concluded that forward exchange policy has great potential as a weapon in the fight against speculative runs on currencies and that the usefulness of the policy as a means of influencing short-term capital movements during normal times depends on empirical judgments, which only actual working in the markets can provide.

This essay was submitted to Yale University in 1962 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics at Stanford University.

Dissertation Committee: Professors Robert Triffin (Chairman), Michael Lovell, and Gustav Ranis.

*

This dissertation was written under the supervision of Robert Triffin and Michael Lovell. Gustav Ranis was the third reader. I wish to thank these gentlemen for their valuable guidance and comments.

The empirical work was made possible through the generosity of the Dresdener Bank, Frankfurt, The Morgan Guaranty Trust Company, New York, and the Bank of France, Paris, which allowed me to use statistical data drawn from their internal books.

I thank the Cowles Foundation for Research in Economics at Yale University, and the Ford Foundation for their financial assistance which made possible the statistical computations and paid for the cost of microfilming the data.

Last, but not least, I acknowledge deep indebtedness to my wife. Her untiring and patient assistance in the preparation of the data, matters of style and typing has been truly invaluable.

H. G. G.

A MULTICOUNTRY MODEL OF FORWARD EXCHANGE: THEORY, POLICY, AND EMPIRICAL EVIDENCE 1955-1961

Herbert G. Grubel

I. Significance of Forward Exchange Policy

1. Introduction

Secretary of the Treasury Douglas Dillon said on June 19, 1961:

> We have, meanwhile, initiated a number of measures designed to diminish the likelihood that speculation against the dollar might recur. Our decision to undertake limited operations in forward exchange markets represents one step in this direction. . . . It is our intention to conduct similar operations in other major currencies whenever such action appears appropriate and useful. . . Although the recent official operations in the forward exchange market have been directed primarily at suppressing potential speculation on currency revaluations, essentially the same techniques can be used to exert an influence, upward or downward, on the covered interest incentive to move short-term investment funds from one market to another. [1]

In the light of the current strong interest in forward exchange policy as a tool of control over currency speculation and short-term capital flows as exemplified by this statement from Secretary Dillon, this dissertation develops a theoretical model of forward exchange rate determination, tests its validity against available empirical evidence and explores the theoretical model's implications for a forward exchange policy. The model builds to a large extent on earlier work done by Keynes, Einzig, Kindleberger, Spraos, Jasay, Tsiang, and Reading. (See Section II for greater detail.) It is felt, however, that a systematic integration of various elements of existing theory makes this model more generally valid than any of the others presented before in literature. An over-all effort is made to integrate forward exchange theory with modern concepts of portfolio choice under uncertainty.

In addition to representing a general synthesis, the model contains several features which have been neglected by earlier writers, but which are important for forward exchange rate policy. These are mainly the influence of triangular arbitrage in forward exchange arising in a more-than-two-country model, the importance of forward exchange rates for actual exports and imports, and the existence of a firm belief by traders in the permanence of spot intervention points. Furthermore, the thesis tries to throw some light on other important questions. For instance, which interest differentials are relevant for the computation of the covered interest incentive? What determines the maturity structure of forward rates? How does the Euro-Dollar market fit into the theory?

---

1. U.S. Congress, Subcommittee of the Joint Economic Committee, Hearings, International Payments Imbalances and Need for Strengthening International Financial Arrangements (Washington: U.S. Government Printing Office, 1961), p. 27.

The empirical work complements the theoretical section by stressing the multilateral aspects of forward exchange and uses data related to five financial centers of the world. It also investigates the validity of theoretical considerations relating to different pairs of interest rates and the maturity structure.

In Section IV, a theory of forward exchange policy is presented which uses the concepts developed in the theoretical model. Special emphasis is put on the relationship between social and financial costs and benefits that can be expected from the use of such a policy.

The present introductory section sets the question of forward exchange policy into the much larger framework of international short-term capital movements and the problems associated with them. This analysis is presented here in order to facilitate a judgment about the usefulness of forward exchange policy and whether it deals with the causes or only the effects of these problems.

2. Hot Money Flows

Much confusion about the subject of international short-term capital movements is caused by the failure to distinguish between two broad types of such capital movements. One type, best called "normal" is characterized by flows in the direction of statistically observable advantages in earnings opportunities. The second type is that of "hot monies" which Nurkse described as "erratic and unresponsive to interest rates, massive, unpredictable and unproductive."[1] Although all large scale short-term capital flows are frequently referred to under this label of "hot money," especially in the financial press, it is advisable to limit the term to those with the characteristic described by Nurkse.[2]

Underlying such hot money flows are often forces outside the sphere of economic analysis, such as political instabilities and mass psychosis. In the past, some of them may have had their root in economic maladjustments, but their significance arose out of the fact that they aggravated what otherwise might have only been a minor disturbance from equilibrium. Thus, they may force the devaluation of a currency through depletion of a country's reserves

---

1. Ragnar Nurkse, International Currency Experience, League of Nations (Geneva, 1944), p. 80.

2. Another classification of short-term capital movements depends on whether they are equilibrating or disequilibrating, i.e., whether they flow into the country with an otherwise unfavorable balance of payments or whether they aggravate the prevailing imbalance. See A. I. Bloomfield, "Postwar Control of International Capital Movements," American Economic Review, 1946, p. 388; C. P. Kindleberger, International Economics (Homewood, 1953), pp. 333-36; S. Enke and V. Salera, International Economics (New York, 1947), p. 153. For the purposes of the present analysis, the alternative classification used is more useful because its focus is on causes rather than effects of short-term capital movements.

between the time when a balance of payments deficit first arises and the time when remedial actions become effective.

Such money flows cannot be prevented with normal tools available to the economic policy maker short of outright controls. Their prevention is largely a matter of politics and the maintenance of an atmosphere of stability and confidence in Western economies. The success of dealing with them effectively after they have occurred depends on how much central banks and their governments cooperate and trust each other's integrity. But this, too, is more a matter of politics than economics.

As will be argued below, the main problems with international short-term capital movements in the future may not be due to such autonomous hot money flows but rather to normal movements of funds seeking a statistically observable higher earning opportunity. The following part will analyze the nature and cause of such "normal," but nevertheless disequilibrating, money flows.

3. International Short-Term Capital Movements in the Classical Model

In the classical gold standard model, private international short-term movements play an important role. When business cycles or random disturbances such as bad harvests or wars cause an excess of imports over exports, the country is assumed to pay for these imports with gold. But, it is not necessary to pay for all of the excess with gold because of the existence of short-term capital flows. These are brought into action when the initial loss of gold causes interest rates to diverge, raising them in the gold-losing country and lowering them in the gold-gaining country.

The size of the trade deficit and the elasticity of supply of short-term funds determine how far interest rates have to diverge. Thus, as the trade deficit develops, some gold leaves the country. However small that gold movement, interest rates in the countries involved tend to change. If the elasticity of supply of short-term funds is great enough, a minute interest differential thus created may attract enough funds so that no further gold movements are required.

But how does this help the adjustment of the basic disturbance which can be achieved in real terms only by a change in price levels or income effects? Doesn't the inflow of funds allow the maintenance of domestic demand and thus prevent adjustment? The answer to these questions is that short-term capital movements do not prevent the adjustment from taking place, they only slow down its progress. Income effects via the multiplier continue to be effective and however small the rise in the interest rate from its initial equilibrium rate caused by the gold loss, this higher rate will reduce aggregate demand somewhat. Eventually prices will fall and a logically equivalent development in the gold-gaining country will bring about the adjustment.

But it is clear that the reduction in demand will be smaller and the price changes slower the less the interest rates move from their original equilibrium position. Therefore, in order to call forth a given reduction in demand, the interest rates have to be maintained at their new level longer,

the smaller the change from equilibrium. Short-term capital movements thus bring the time dimension into the adjustment process in an essential way.

In a world where adjustments are not instantaneous, the time element is important. A lengthening of the adjustment process means fewer and less severe disruptions of the ordinary course of business because prices and wages are more flexible in the long run; unemployment and misallocation of resources are reduced. In general, then, it can be said that in the classical model international short-term capital movements serve as a beneficial cushion which protects the domestic economy from excessive stresses and strains caused otherwise by disturbances originating in foreign countries. They also reduce the need for gold and reserves in each country. What then explains the current concern with international money flows and the search for methods to control them?

4. Real World Imperfections and Conflict of Policies

There are some important ways in which the real world differs from the theoretical model just presented, and it is these differences which lie at the bottom of today's difficulties with international short-term capital movements. For one, there has probably never been a direct and complete relationship between gold losses and interest rates, even in the absence of conscious national interest policies. Cyclical or institutionally determined variations in the velocity of circulation of money may have stronger influences on rates than changes in the quantity of money, leaving us with a mere tendency of interest rates to shift in the proper direction. Most important of all, however, is one characteristic of prices and wages which the classical model fails to take into account. This is that they are rigid in a downward direction to a degree which makes efforts at lowering them exceedingly costly in terms of unemployment and growth. This factor is probably most to blame for the reluctance of many countries to accept domestic price changes when one trading partner experiences some unique shock of cyclical disturbance to internal stability. This is especially so because external stability requires, with certainty, a subsequent readjustment of the price level. And since the quantity of short-term flows is determined solely by the market, these adjustments may have to be very rapid.

Countries today are unwilling to have external events and unmitigated market forces exert much pressures on domestic conditions. Since the end of World War II, it has become the declared national economic policy objective of most Western countries to maintain full employment of resources and price stability. The traditional and most widely used tool with which these objectives are achieved is monetary policy and changes in interest rates. If each country were to set its interest rate at the level ensuring a maximum degrees of internal stability, then interest differentials between countries would be uniquely determined.

Given these interst differentials between countries, the elasticity of supply of international short-term funds determines the size of the capital movements just as in the classical model, with freely adjusting interest rates,

this elasticity determined how great the differential would be. And here lies the conflict: it is impossible to set both the interest differential and the size of the flows. But, as long as countries are preoccupied with domestic stability, they will continue to set interest rates for this purpose. And this results in the problem of international short-term capital flows.[1]

The elasticity of the supply of these funds is crucial in determining the size of the flows and the magnitude of the problem. Rather inelastic supply schedules are not too burdensome. It means that countries will forego the benefits of equilibrating flows and need to meet balance of trade deficits with previously accumulated reserves. More burdensome, however, is the situation in which the supply of funds is so elastic that even small interest differentials result in massive shifts of funds which tend to tax the existing reserves. The world has been moving rapidly from the first situation to the second. Immediately after World War II, exchange restrictions prevented effective interest arbitrage. But, with the re-establishment of convertibility at the end of 1958, these restrictions were removed. The expansion of international trade, increasing confidence in the stability of Western economies, improved methods of communications and warfare techniques (making Paris as safe as New York), and the growth of world wide manufacturing and banking enterprises have been important contributing factors to the increase in the elasticities of the supply of internationally mobile funds. These developments can be expected to continue in the future and may make this problem of the conflict between internal and external balance one of the greatest facing the Western World in the next decade.[2]

5. Forward Exchange Policy as One Solution to the Dilemma

The analysis has shown that national interest rate policies and business cycles give rise to the problems of international money flows which above have been called "normal." Forward exchange policy can be used to greatest advantage in the handling of these, as will be seen in Section IV. But, there are also implications for the control of speculative attacks on currencies. The

---

1. Currently, various economic policies are being tested and discussed, such as a greater reliance on fiscal policy or attempts at changing the shape of the yield curve of securities. These are aimed at reducing the dilemma associated with the use of interest rate policies. These tools will be listed later, but detailed analysis of them goes beyond the scope of this study. Instead, focus will be on forward exchange policy, the potential of which has been the initiating force of this study, and which is just one of the methods of dealing with the problem just analyzed.

2. "The most important economic problems of the decade may well be . . . the limitation on massive international shifts of short-term funds." T. O. Yntema, "Economic Adjustments Among Nations," The Journal of Finance, XVI (March 1961), p. 5.

Hearings before the Subcommittee of the Joint Economic Committee, loc. cit., see especially the testimony by Bernstein, Dillon, Hayes, Heller, et al.

problems of arbitrage and speculation are closely intertwined in the sense that large scale arbitrage depletes reserves which can set off waves of speculation. Control of arbitrage may thus bring a measure of control over speculation. This is especially important if countries learn to avoid political and balance of trade situations that in the past have often triggered speculative runs.

Many solutions to the problem of international short-term capital movements have been proposed. They fall into two groups. One is centered around giving up interest rate policy as a tool for the achievement of domestic stability. In so far as this solution means giving up influence over the economy it is not likely to be acceptable to politicians and policy makers at a time of challenge from the rapid economic progress in Communist countries, increasing defense and social needs which require stability and sustained growth in the Western World. Because of this, the proposals involving a diminished role for interest rate changes are usually coupled with recommendations for the use of other methods with which the influence over the economy can be retained. The most important of these is, of course, the recommendation to use fiscal policy more readily and flexibly than in the past in order to achieve the desired freedom for interest rates. [1]

Partly belonging in this group and partly belonging in the next group of solutions is the effort to change the yield curve on securities. This involves changing long-term interest rates by open market operations to suit domestic needs, while at the same time trying to set short interest rates in order to minimize disturbing international short-term capital flows. [2]

The second group of solutions aims at retaining monetary and interest policy in its traditional use. It envisages combatting the impacts of the resulting international money flows. The most important of these is the creation of sufficient international reserves (Bernstein and Triffin plans) [3] which would enable countries to weather the large losses in the wake of capital flows, or simply the provision of stand-by agreements between countries to help each other in case of emergency (discussed at the IMF meetings at Vienna in September 1961).

All of these methods and others not mentioned are being considered by economists and politicians. They each have certain problems, advantages and disadvantages associated with their adoption. Forward exchange policy is one solution that has been recommended as a complement to the others. It has been claimed that this policy will contribute materially to a solution of the basic

---

1. Money and Credit, Report of the Commission on Money and Credit (Englewood: Prentice Hall, 1961), p. 247.

2. Ibid. See also, The Economic Report of the President, the Annual Report of the Council of Economic Advisers (Washington, January 1962).

3. E. M. Bernstein, "The Adequacy of United States Gold Reserves," American Economic Review, LI, Papers and Proceedings (May 1961), p. 439.
R. Triffin, Gold and the Dollar Crisis (New Haven: Yale University Press, 1960).

dilemma of one interest rate influencing both internal and external stability, as well as help reduce the impact of speculative attacks on currencies. The following three sections of this study are aimed at exploring the validity of these claims.

II. Toward a General Theory of Forward Exchange

1. Introduction

In this second section, the pure theory of forward exchange is presented and the forces which guide the actions of the participants in the forward exchange markets are analyzed. The resulting model serves the dual purpose of allowing (1) the evaluation of the empirical evidence, and (2) the critical analysis of the question whether official intervention in the forward exchange markets can be used to control the movements of private short-term capital.

The section is divided into six parts. The first is a glossary of terms used in the discussion of the forward exchange markets. The second describes one group of participants in the markets lumped together under the name of arbitragers. The third part deals with the second group of market participants, loosely called speculators-traders. In part four, the interaction of speculators-traders and arbitragers is explored in order to explain how forward rates are determined. In part five, the model is modified to account for the existence of spot-intervention points. In part six, a theory of forward exchange maturity structure is presented.

2. Glossary of Terms

The analysis of forward exchange markets is not difficult in the same sense in which Einstein's theory of relativity, or capital theory is difficult. They demand highly abstract reasoning. Most difficulties in connection with the theory of forward exchange markets arise out of the fact that the theory involves a great deal of what has been called "mental gymnastics." The following definitions of terms [1] are aimed at reducing the troubles associated with these mental gymnastics by giving a precise definition of the basic concepts whose numerous interconnections are to be analyzed in subsequent parts.

Spot-rate—The cost of one unit of foreign currency in terms of domestic currency for delivery on the following day. Symbol X_0. Example: in New York, the price of one £ is quoted as \$2.80, i.e. $(\$ \cdot £)_0 = 2.80$.

Forward-rate—The cost of one unit of foreign currency in terms of domestic currency for delivery at a stipulated time from the date of the contract. Symbol X_t, where subscript t refers to the time involved between the day the

1. The terms used are those introduced by the following:
P. Einzig, The Theory of Forward Exchange (London, 1937), and A Dynamic Theory of Forward Exchange (London, 1961).
B. Reading, "The Forward Pound 1951–59," The Economic Journal, LXX (June 1960), p. 304.
S. C. Tsiang, "The Theory of Forward Exchange," International Monetary Fund Staff Papers, VIII, No. 1 (April 1959), p. 76.

contract is made and the day the forward currency is to be delivered. "t" may be from one day to several years, but market rates are usually quoted for 30 and 90 days.

Discount—From the New York viewpoint, sterling (i.e., the foreign currency) is said to sell at a forward discount whenever $X_t < X_o$. This implies that the forward dollar (domestic currency) is at a premium.

Premium—By analogy, the foreign currency sells at a forward premium whenever $X_t > X_o$. This implies the domestic forward currency is at a discount.[1]

Implicit rate of interest—Premiums and discounts are often expressed as percentage earnings per year:

$$IR = (X_t - X_o)/X_o \cdot (360/T)$$

This means that the implicit rate of interest is the difference between the spot and the forward rate as a percent of the spot rate multiplied by the number of times the length of the contract would allow the repetition of the transaction within a year. Read: $(X_t - X_o)/X_o \cdot (360/T) < 0$ as a discount; $(X_t - X_o)/X_o \cdot (360/T) > 0$ as a premium; $(X_t - X_o)/X_o = 0$ as par.

Interest rate differential—This is the difference between domestic and foreign interest or discount rates in percent per year on "equivalent" securities. Symbols $ID = R_d - R_f$. Read: $R_d - R_f > 0$ as interest rate differential in favor of domestic market; $R_d - R_f < 0$ in favor of foreign market; $R_d - R_f = 0$ as parity.

Intrinsic interest rate differential or covered interest arbitrage margin—This is defined as $ID - IR$. Whenever the advantage a foreign money market possesses in terms of a simple interest rate differential is not completely wiped out by the cost which the implicit interest rate represents, a positive intrinsic interest rate differential (or covered interest arbitrage margin) exists. When X_t is such as to make $ID - IR = 0$, X_t is considered to be a parity and is called X_t^*

Covered interest arbitrage—Anyone who invests in a foreign market while at

1. To illustrate the mental gymnastics involved in this subject and to warn the reader of this source of confusion which may arise from reading British writings, the following differences between practices in London and all other exchange markets is presented.

London quotes exchange rates in terms of how many foreign units of exchange are required to purchase one pound sterling. The above mentioned rule as to when a domestic currency is traded at a discount or premium therefore must be reversed. As an illustration, consider the sterling-dollar rates quoted in London: spot = 2.80, forward = 2.82. According to the above definition $X_t > X_o$, and therefore we would say that the foreign currency in London sells at a premium. But, given these exchange rates, the foreign currency is at a discount. This can be seen by "normalizing" these rates and expressing them in terms of how many domestic units it takes to purchase one dollar. This is done by making spot = 1/2.80 and forward = 1/2.82. The result is $X_t < X_o$. The foreign currency is at a discount. Subsequently, whenever the term premium or discount are used, they refer to the usage employed outside of London.

the same time eliminating the exchange risk by selling the foreign currency forward is considered to engage in covered interest arbitrage.
Triangular arbitrage—This action involves taking riskless advantage of an inequality in the following expression: $(\$ \cdot £) \lesseqgtr (\$ \cdot DM) * (DM \cdot £)$ by buying and selling the currencies involved. The expression tends toward equality regardless of whether the three rates are the spot or the forward rates.
Speculation—This is the act of deliberately exposing oneself to additional risk of exchange uncertainty.[1] A "long position" is taken by purchasing foreign currency forward without owing a spot payment in the future in the expectation that the spot sale of the foreign currency received at the specified time in the future will produce a profit. A "short position" involves the forward sale of a foreign currency without current ownership of an equivalent amount of this currency, again with the expectation that at the specified future time the spot rate will be such as to make the transaction profitable.
Hedging—This term describes the act of eliminating an exchange risk due to: (a) normal international commercial transactions, or (b) foreign investment of short-term capital funds as part of the covered interest arbitrage operation by the purchase or sale of forward currency.

3. Arbitragers

The current part is devoted to the analysis of a group of participants in the forward exchange markets who have the common characteristic of always eliminating any exchange risks caused by their activity. The analysis assumes that government regulations interfere with neither the arbitragers' basic activity which gives rise to the exchange risk nor with the act of eliminating the risk. For purposes of exposition, it is convenient to assume at first the existence of only two countries and demonstrate basic relationships in this framework. In the second part, however, the analysis is expanded to include more than two countries. It is assumed, furthermore, that exchange rates are flexible without limit. The removal of this assumption is undertaken in part six of this section. Basic to the theory of forward exchange are the interest arbitragers who are made up of two groups: (a) owners of funds, and (b) borrowers.

(3.a) Owner Arbitragers

Individuals, manufacturing corporations, banks or institutional investors often accumulate liquid funds which are earmarked for expenditure at some point in the near future or serve as a liquidity buffer for unforeseen expenditures. For example, dividend or interest payments on capital or tax liabilities are due on specific dates in large sums and have to be accumulated over time. Capital raised through new issues often accumulates in lump sums and is deployed only gradually on real expenditures. Investments in the relatively capital-certain money market instruments such as treasury bills, bankers acceptances, trade or finance paper provide these investors with the opportunity to retain ready

1. This definition is used by Tsiang, loc. cit., and Kindleberger, loc. cit., p. 40.

access to their funds but yet allow the earning of some interest. Competing for the funds of investors with these domestic instruments are foreign money market papers. The basic considerations which determine the investors' choice of domestic or foreign assets are as follows.

An owner of short term funds in New York will be willing to move them to London if the return he can earn in New York (R_d for three months) is smaller than that in London (R_f for three months) after going through the process of buying sterling (at X_o and paying a commission equal to E%) with which he purchases the London securities and then at the same time eliminating the exchange risk by selling the sterling forward (at X_t, T for 90 days and again paying a commission of E%). An inequality makes this condition precise.[1] Funds will flow from New York to London when:

$$(1 + R_d) < (1/X_o)(1-E)(1 + R_f)X_t(1 - E)$$

Disregarding transactions costs, which may in fact be quite negligible, and simplifying the expression, it can be written as:

$$1 + R_d < (X_t/X_o)(1 + R_f) \qquad (1)$$

When the inequality is reversed, funds will flow from London to New York:

$$1 + R_d > (X_t/X_o)(1 + R_f) \qquad (2)$$

In equilibrium, when the two sides of the expression are equal, no funds will flow:

$$1 + R_d = (X_t/X_o)(1 + R_f) \qquad (3)$$

Traditionally, this equilibrium condition is said to be equivalent to the statement that $R_d - R_f = (X_t - X_o)/X_o$; i.e., the interest rate differential is is equal to the implicit interest rate. The difference between the two sides of the equation has been called the "crude arbitrage margin." Algebraic transformation of the more basic equilibrium statement, however, produces the "precise arbitrage margin": (PM)[2]

$$PM = (R_d - R_f)/(1 + R_f) - (X_t - X_o)/X_o$$

---

1. This algebraic presentation was adapted from J. Spraos, "The Theory of Forward Exchange and Recent Practice," Manchester School of Economic and Social Studies, XXI (May 1953), p. 88. The continuous case is presented by E. Sohmen, Flexible Exchange Rates (Chicago University Press, 1961), pp. 66-69.

2. Spraos, op. cit., p. 88-89.

If R_f is large, the difference between the crude and precise margins may be significant. [1]

(3.b) Borrower Arbitragers

Besides the owners of investible funds, there is another group of participants in the market who are motivated by the same considerations summarized in the concept of the covered arbitrage margins. This group is made up of borrowers. This act of borrowing spot funds and covering the exchange risk is to be separated analytically from the simple purchase of forward currency to guarantee the exact revenue in domestic currency from a commercial transaction (see part 4). The distinction of these two operations is not always clear in the literature and is important because one operation involves spot funds and therefore a country's reserves while the second does not. [2]

Usually these borrowers are business men in need of financing their trade. What is involved can be demonstrated most easily by the use of an example. A New York importer purchases a stock of English goods. If his supplier wants immediate payment of the value in sterling, the importer may have to borrow the money to meet this demand. Let us assume he intends to repay the sum after the goods' arrival and sale in New York three months later. He can take out a loan either in London in which case he owes sterling three months later, or he can take out the loan in New York where, by buying spot sterling, he settles his merchandise debt and ends up owing dollars. If interest cost were the only consideration, he would borrow at the place with the lowest rate. But there exists also an exchange risk which arises out of choosing the first alternative because the trader cannot know how many dollars it may take to repay his sterling loan three months from now. Purchasing forward sterling, however, provides him with the opportunity to eliminate this risk. Which of

---

1. In empirical work, two different approaches are possible: (1) the R_d and R_f usually quoted in rates per year can each be divided by four to make them comparable with the 3 month implicit interest rate $(X_t - X_o)/X_o$ which is left uncorrected. Or, (2) the R_d and R_f are left unchanged and $(X_t - X_o)/X_o$ is multiplied by four (360/90). In the first case, the difference between crude and precise arbitrage margins is smaller than in the second. Published empirical work that has come to my attention relies exclusively on the second method and uses the crude margin. See, Tsiang, op. cit., p. 80, footnote 2, on the theoretical size of error between the crude and precise margins. Use of an electronic computer made it practical to compute the precise margin in the empirical work presented in Section III.

2. Spraos, op. cit., p. 92.
J. Spraos, "Speculation, Arbitrage and Sterling," The Economic Journal, LXIX (March 1959), p. 4.
P. Einzig, The Theory of Forward Exchange (London: 1937), pp. 174-75; 188-89; 217-18.
Reading, op. cit., p. 307.

the two available alternatives he chooses depends on which is cheaper: paying the New York interest rate and current spot rate, or paying the London interest rate and purchasing forward sterling at the current rate.[1]

(3.c) The Euro-Dollar Market

During the past five years, a new institutional arrangement has brought the money markets of the world closer together.[2] It is based on the following considerations: an Italian exporter earns dollars which are on deposit in a New York bank. If he wants to make a temporary money market investment he can do so by purchasing foreign assets in New York using his dollar balances there and covering himself against the exchange risk by the forward sale of dollars. Alternatively, he can sell the dollars spot and invest in the domestic market. The Euro-Dollar market opens a third possibility to him. He can lend out the dollars to a domestic intermediary who may pay him a higher interest than he can get in New York. Repayment of the loan by the intermediary can be in dollars or local currency. If the repayment is in dollars, the owner covers himself in the forward market. A favorable constellation of spot, forward and interest rates can thus produce a return superior to the one to be based on domestic or direct foreign investment.

The intermediary who has gained command over these dollars can use them for his own business as means for direct settlement or can lend them out to a customer. If he lends them to a customer, he more likely than not will exchange the dollars against local currency and cover the exchange risk through forward purchase of dollars. He can, of course, also use them to finance imports from the U.S. For the final user of the dollars as well as the original lender and the intermediary the deal is advantageous, or it would not have taken place. The net advantage is to a great part due to the fact that the final

---

1. The following reasoning in terms of the symbols can be used to show the equivalence of the action of borrower arbitragers with that of the fund owning arbitragers.

The borrower arbitrager is indifferent between London and New York if the sterling obligation of N, times the spot rate X_s, borrowed at the New York interest rate R_d is equal to the cost of interest at R_f on N sterling made into a certain dollar obligation by purchase of forward sterling at X_t. $N(X_o)(1 + R_d) = N(1 + R_f)X_t$, which can be written as $1 + R_d = (1 + R_f)(X_t/X_o)$. In the case of an inequality like this: $1 + R_d > (1 + R_f)(X_t/X_o)$, the fund owning arbitragers would move their funds to New York, the borrower arbitragers would borrow in London. Both of these actions have equivalent effects on the exchange rates, interest rates and reserves of the countries. Borrowing delays the remittance of value for exports and is one way in which the well known "lags" in the balance of payments are brought about.

2. O. Altman, "Foreign Markets for Dollars, Sterling, and Other Currencies," International Monetary Fund Staff Papers, IX (December 1961), p. 313-52.

user in our example can get dollars (or domestic funds) cheaper in this way because his credit rating is better with the intermediary with whom he may be doing regular business than it would be in New York where he might otherwise be forced to borrow. He therefore pays the intermediary a lower rate of interest than would be required in New York. Yet, the rate may be, and most often is, high enough to allow the intermediary to pay a higher interest to the original supplier of the dollars than the direct New York money market investment might have brought. This is so because the intermediary requires a smaller risk premium on the loan, or his overhead may be smaller than that of his New York counterpart (i.e., the difference between lending and borrowing rates is larger in the U.S. than in Europe).

The development of this market for loans in dollars and other currencies has tied world money markets closer together and tends to push forward rates toward parity. For only if the yields from New York investments are lower than the return earned by lending out dollars directly, properly hedged against the exchange risk, will the funds be placed outside New York. This tends to raise interest rates in New York, lower them outside and exert pressure on the forward rate until the advantage of one market over the other tends to be wiped out.

(3.d) Triangular Arbitrage

The third group of arbitragers will be treated next. Their influence toward the establishment of equilibrium in the sense of equation (3) has so far been neglected in the literature.[1] This neglect may have been due to the practice of limiting the analysis to two-country models. The following model will incorporate three countries, but the principles developed will be equally valid for any number of countries.

It is a well known fact that triangular arbitrage creates consistency in spot rates at any given time. The equilibrium condition describing this consistency is:

$$(\$ \cdot DM)_0 = (\$ \cdot £)_0 * (£ \cdot DM)_0 \qquad (4)$$

Its achievement is almost always perfect because of the ease with which profits can be made by the arbitragers neither needing capital nor facing any risks. Transactions costs to the arbitrager are negligible as compared with the possible size of the arbitrage operations. If we assume spot rates to be consistent and take domestic interest rates in each center as given, then according to

1. J. Spraos, "The Theory of Forward Exchange, Recent Practice," Manchester School of Economics and Social Studies, XXI (May 1953), pp. 90-91, was the first writer to incorporate the action of these arbitragers into the theory. But he has failed to analyze the implication which these actions have for forward rate policy.

P. Einzig, Dynamic Theory of Forward Exchange (London 1961), p. 259, examines the effect of triangular arbitrage in the London market alone. The demonstration of consistency in the text was taken from Spraos.

equation (3), in equilibrium the three forward rates are uniquely determined. Will these emerging three forward rates lend themselves to triangular arbitrage or are they consistent? Following is the proof that they are consistent. Equilibrium conditions in each market according to equation (3) are:

$$(\$ \cdot DM)_o = (\$ \cdot DM)_t * (1 + R_F)/(1 + R_{NY}) \quad (5)$$

$$(\$ \cdot £)_o = (\$ \cdot £)_t * (1 + R_L)/(1 + R_{NY}) \quad (6)$$

$$(£ \cdot DM)_o = (£ \cdot DM)_t * (1 + R_F)/(1 + R_L) \quad (7)$$

Substituting in (4):

$(\$ \cdot DM)_t * (1 + R_{NY})/(1 + R_F)(\$ \cdot £)_t * (1 + R_{NY})/(1 + R_L) * (£ \cdot DM)_t(1 + R_L)/(1 + R_F)$ which reduces to:

$$(\$ \cdot DM)_t = (\$ \cdot £)_t * (£ \cdot DM)_t \quad (8)$$

From this reasoning it follows that whenever in a three country model there exists a covered arbitrage margin between one pair of countries while there exists none between the remaining two pairs of countries, then arbitragers using the triangle of inconsistent forward rates will exert strong equilibrating forces which, together with interest arbitragers, tend to eliminate the covered arbitrage margin. The importance of these triangular arbitrage operations arises out of the fact that actual movements of spot funds affecting currency reserves may not be involved.

Equation (8) can be used to point out another fact, important for policy. If both forward rates involving the dollar change in the same direction and by the same percentage, then the equality is preserved and opportunities for triangular arbitrage do not exist, even though the system is in disequilibrium with respect to equations (5) and (6).

(3.e) Qualifications

Until now the argument implied that the forward rates do all the adjusting when an arbitrage margin appears. In reality, however, the spot rate, interest rates in both centers, and the forward rate are determined simultaneously. The investor moving his funds from New York to London adds to the demand for spot sterling, tending to drive its price up. The New York market for investible funds feels an upward, and the London market a downward, pressure on interest rates as a result of the arbitrage operation. It is important to be aware of these effects. For the purposes of analysis, however, it is convenient to focus attention on the determination of the forward rates alone. Assumptions making this procedure plausible are that the size of the funds involved in covered interest arbitrage operations may be insignificant as compared with the other determinants of interest and spot rates, or that the government pegs these rates.

(3.f) The Arbitrage Schedule—The Diagram

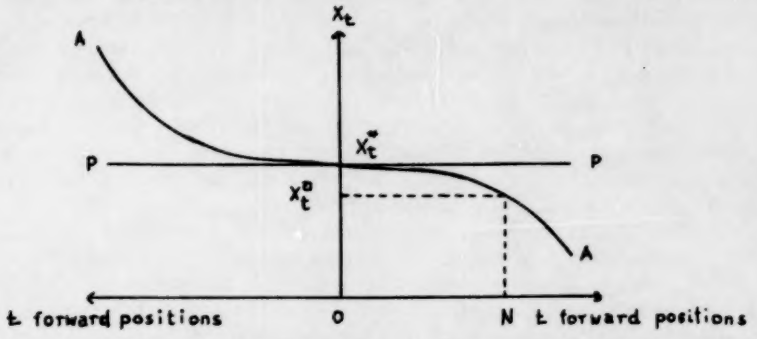
To make the relationship between the size of forward currency

commitments and the forward exchange rate more explicit, use will be made of a diagram.[1]

At first, only the demand for currency by covered interest arbitragers is analyzed, for each dollar of foreign portfolio investment gives rise to a dollar's purchase of foreign forward currency. Because of this one to one relationship, the focus of the analysis can be on the investment borrowing behavior of individuals when an explanation of the shape of the demand curve for forward exchange is attempted.

Fig. 1

Arbitrage Schedule



In Figure (1) the vertical axis measures spot and forward rates. Along the horizontal axis to the right of zero, the stock of forward sterling bought by investors is measured; to the left of zero, forward commitments to deliver sterling are shown. According to the above reasoning, when the spot rate and the interest rate differential are given, the parity forward rate (X_t^*) is known. Through it the horizontal line PP is drawn. The arbitrage schedule AA indicates forward currency positions from arbitragers to be zero at $X_t^* = X_t$ and then to be an increasing function of the absolute difference of $X_t - X_t^*$. Thus a forward rate of $X_t = X_t^a$ brings forth net forward commitments from interest arbitragers equal to N. At some quantity, the AA schedule becomes vertical.

(3.g) Choice of Interest Rate Differential

Before turning to an explanation of the shape of the AA schedule, another matter must be discussed. Money markets have developed many kinds of securities with equal maturities, all of which carry different interest rates, such as bankers' acceptances, trade bills, treasury bills, finance company paper, savings deposits, etc. Between two countries there exist, therefore, more than one interest differential even for securities with the same maturity. Which one of these differentials will be the basis on which the parity forward rate (X_t^*) is

1. The diagrammatic exposition owes much to E. A. Jasay, "Bank Rate or Forward Policy," Banca Nazionale del Lavoro Quarterly Review, 44 (March 1958), pp. 72-73, and B. Reading, op. cit.

computed and policy decisions made? Some light can be shed on this question through an empirical investigation as described in Section IV. From a theoretical point of view, the following answers suggest themselves.

The most simple of these is that it does not matter which interest rate differential is chosen since they are all equal. Assumptions making this plausible are that bankers' acceptances always have a yield of .005% above that of treasury bills in New York as well as in London. [1] Under these assumptions, the interest differential between any pair of securities would be the same regardless of the absolute size of the basic interest rates. If, however, the yield on bankers' acceptances is typically above that of the Treasury Bills by a margin of a certain percent of the yield itself, the differentials would depend on the absolute levels of the rate on bills and would not always be the same even though percentage differentials between assets are the same in New York and London and changes are closely correlated. Under such circumstances and others that produce differentials of unequal size between securities of equal maturity, the parity forward rate can be thought of as applying theoretically to a weighted average of such rate differentials.

A complicating factor in this analysis may be that investors prefer assets of one quality in one market and assets of a different quality in another market. This possibility cannot be ruled out. [2] But such behavior is not very rational and can be expected to decline in importance as information about foreign markets becomes more readily available. In general, the bias introduced by the use of the yield differential between Treasury Bills by the above considerations is likely to be of a second order of smallness because the risk premia separating the yields on typical money market securities of equal maturity are rather small. Thus, with Treasury Bill rates in two countries equal at 2.0%, the switch into bankers' prime acceptances in the foreign market yielding typically 2.2% when Treasury Bills bring 2.0% would bring an advantage of only .2%. But again the question arises why the investor would not choose to purchase domestic bankers' acceptances if he is willing to purchase them abroad.

---

1. An unpublished study of the correlation between changes in interest rates of short term securities with equal maturity but different risks carried out for the U.S. by A. Okun of the Cowles Foundation of Yale University shows that the correlation is very high. But it is clear that such a correlation does not imply constancy of the interest differentials.

2. The investment officer of a large institution whom I interviewed about his awareness of foreign short term investment opportunities was willing to compare only the yield on money market paper issued by U.S. finance companies (in which he regularly invested) with U.K. Treasury Bills. The covered yield margin of U.K. commercial paper over U.S. commercial paper did not interest him because of the lack of information about the soundness of the U.K. debtors.

(3.h) Portfolio Choice and Foreign Assets

Returning to the main argument, it must be pointed out that the AA schedule drawn in Figure (1) is different from that found in the literature. It does not exhibit stretches of perfect elasticity and kinks at two points as the one used by Reading and Jasay. Underlying these kinks is the concept of a critical value at which it becomes profitable for a very large number of investors to move their funds. At rates lower than the critical value (1/2%) only very sensitive investors such as foreign exchange departments of banks are assumed to take advantage of higher earnings in the foreign center. At another point, it is assumed that the most mobile funds having moved, greater earnings than 1/2% have to be offered to call forth further shifts. This produces another kink.

The shape of the smooth AA schedule shown in Figure (1) can be defended and the traditional one criticized on two grounds. One is Einzig's report[1] of his interviews with London bankers and investors who assured him that no such critical value exists in fact and that they are willing to move funds into foreign centers in response to very small arbitrage margins, as low as 1/10 of one percent.

The second ground is that investors in the real world rarely behave as though their demand schedules had kinks which means that they would be switching completely from holding one type of asset to another at some critical yield differential.[2] The theory of portfolio choice under uncertainty as developed by Tobin[3] and Markowitz[4] can be used to explain how a greater arbitrage margin causes an increased holding of foreign assets and forward exchange in a smooth and continuous way. In the following, the basic Tobin model is presented and applied to foreign assets covered against the exchange risk.

When individuals decide on the proportion in which they want to hold various types of assets in their portfolios, they usually are not absolutely certain what the return from any given asset will be. Instead, they have in mind a number of possible outcomes, each carrying a certain probability of occurring. The mean of such a probability function associated with each asset is called the expected rate of return (E(R) or μ). The standard deviation

---

1. P. Einzig, "Some Recent Changes in Forward Exchange Practices," The Economic Journal, LXX (September 1960), p. 486.

2. On this question of one-asset portfolios, implicit in traditional analysis, see J. Tobin, "Liquidity Preference as Behavior Towards Risk," Review of Economic Studies, XXVI (February 1958). This article also describes how theoretically smooth aggregate demand schedules can be derived from individual demand schedules exhibiting kinks as long as the critical values are different for each individual.

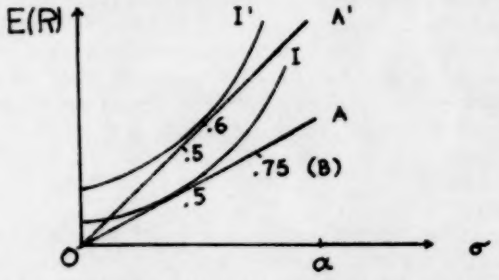
3. Ibid.

4. H. Markowitz, Portfolio Selection, Cowles Foundation Monograph 16.

of the function is a convenient measure for the riskiness of the asset. The investor maximizes his expected utility by choosing among a menu of assets as suits his personal preference. The accompanying diagram (2) illustrates such a choice.

Fig. 2

Model of Portfolio Choice



Along the vertical axis E(R), and along the horizontal axis σ are measured. The investor is assumed to face a choice of only two assets, cash with E(R) = 0 and σ = 0 and a foreign asset with E(R) = a and $\sigma = \alpha$. The line OA is the opportunity locus of combinations of E(R) and σ open to the investor. He can attain these different combinations by varying the proportion of foreign assets in his portfolio. This proportion is one at point A; .75 at point B. The indifference curve I is tangent at the point where the proportion of foreign assets and cash in the portfolio is .5. The shape of the indifference curve is an essential part of the argument. Its derivation and a justification for its use can be found in Tobin's article.

The E(R) on foreign assets is an increasing function of the arbitrage margin. If we assume that the risk associated with the asset is independent of the size of the arbitrage margin, then a higher E(R) on foreign assets caused by a widening of the differential between X_t^* and X_t results in a new opportunity locus OA'. The higher indifference curve I' is tangent at a point which indicates that the higher return on the foreign asset caused a substitution of the foreign assets for cash so that now the proportion is .6. Thus the higher return on the foreign asset causes the demand for it to be increased. The same principle which applies to owner arbitragers also applies to borrower arbitragers.

The analysis involving two assets with nonzero risk is more complicated but its conclusions remain essentially the same. Foreign assets associated with a greater risk than their domestic counterpart compete for a place in the portfolio just as any other asset would, which had the same riskiness and expected rate of return.

As demonstrated mathematically in an appendix to the dissertation proper the substitution of a given amount of cash for foreign assets requires an increasingly larger arbitrage margin as the proportion of cash held approaches zero.[1] This property of the functions underlying the model explains

1. The demonstration in the appendix is limited to the case of a two-asset portfolio where one asset yields no return. For the more complicated and

the inelastic parts of the AA schedule. The AA schedule is vertical at the limit set by the net worth of investors and the credit lines of borrowers.[1] The vertical part of the AA schedule may occur sooner when some investors are prohibited by organizational by-laws from investing more than a certain proportion of their portfolio in foreign assets.

(3.i) Maturity of Forward Contracts

As mentioned above, looking at owner-arbitragers alone the relationship between forward contracts outstanding and the stock of short term funds in the foreign center is one to one. This is so regardless of the length of the investment or the length of the forward contract. To illustrate this, consider the stock of $1 billion in London having moved there in response to a covered margin of .6%. Looking at the figures at any given point in time during a year when the margin remained unchanged, $1 billion of forward contracts was outstanding. This is obvious in the case where the forward contract is for a year. But if the initial investment had been in one month bills, covered by one month forward sales, the case is not so obvious. What happens is that at the end of each month, the investor has to go through the same considerations which underlie his initial shift and decide where to keep his funds. If the covered earnings differential persists, he will keep them in London. Logically, he then goes through the following steps: receipt of funds from maturing investment, exchange of the funds for home currency on the basis of the old forward contract, purchase of spot currency, reinvestment and cover through new forward contract.[2] This process is in practice short circuited in that deliveries of funds in the exchange dealings are not made but the forward contract is renewed and the difference between spot and forward is settled in cash.[3]

---

realistic situation where both assets carry a positive expected rate of return and risk, the following intuitive explanation is offered.

An investor with an original portfolio of $500 asset A, yield $\underline{a}$, and $500 asset B, yield $\underline{b}$, may require a yield $\underline{c}$ on asset B to induce him to hold another $100 of B, where $c > b$. To have him add still another $100 of B after this is likely to require an even higher yield $\underline{d}$, where $d > c$.

1. J. M. Keynes, Monetary Reform (New York 1924), p. 140, ". . . the floating capital normally available, and ready to move from center to center for the purpose of taking advantage of moderate arbitrage profits between spot and forward exchange, is by no means unlimited, and is not always adequate to the market's requirements."

2. This analysis supports Auten's rejection of the view expressed by Tsiang, op. cit., p. 102, that pegging may lead to a cumulatively growing quantity of forward commitments. For Auten's views, see "Counter Speculation and the Forward Exchange Market," Journal of Political Economy (February 1961), pp. 49-55.

3. Tsiang, op. cit., p. 89, footnote 10.

(3.j) The Dynamics of Interest Arbitrage

The diagram and the underlying analysis are static. They tell nothing about the actions of arbitragers between equilibrium positions, i.e., how they go about adjusting their portfolios when the arbitrage margin changes. For this, dynamic analysis is required. The problems involved in such an analysis are formidable and the following assumptions are introduced in order to make the problem manageable:

(1) All foreign investments occur in three month bills and are covered in the three month forward market.

(2) There exist no facilities for odd day forward contracts and all investments have to be held to maturity.

(3) Arbitrage margins are caused by the change of the forward rate and the implicit rate of return (i.e., $(X_t - X_o)/X_o$). Interest rates and spot rates in the two countries do not change.

(4) The maturity structure of his portfolio does not concern the investor.

Under these assumptions, when the forward rate changes such that the schedule calls for an increase in forward exchange holdings, the adjustment occurs by investors' purchases of additional foreign investment which they cover in the forward market. If the change of the forward rate requires a reduction of forward commitments by arbitragers, this can be achieved only by letting old contracts mature and not renewing them which means the domestic reinvestment of the sum.

If the above assumptions are dropped, the investor has to consider so many variables in deciding by what method to adjust his foreign asset and forward exchange holdings that generalizations are not possible. All that can be done here is to enumerate some of the considerations an investor has to go through in deciding how and with what speed to adjust his foreign asset holdings. When the expected return on foreign assets is raised, the required increase in holdings is accomplished simply by the purchase of assets with the desired maturity. The matter is more complicated when the change in the expected return calls for reduced holdings of foreign assets.

There may or may not be a capital gain or loss from the sale of the security itself. Similarly, the undoing of the forward contract may result in a capital gain or loss which can either mitigate or accentuate the result of the security disposal. The investor always has, of course, the choice to let his asset mature and earn fully the rate originally aimed for. Entering this choice may be expectations of what the rates will be at the date of maturity. All of these considerations are influenced by the way in which the reduction in the arbitrage margin has been brought about: change in forward rate, spot rate, domestic or foreign interest rates, each alone or in some combination. While simple statements about the adjustment through time of interest arbitragers in response to changes in the arbitrage margin are not possible, once all parameters relevant to the decision are specified, it is possible to determine what actions a rational and profit-maximizing arbitrager is likely to undertake.

(3.k) Triangular Arbitragers

The shape of the AA schedule shown in Figure (1) is not modified in an

essential way by the inclusion of another group of arbitragers which above has been called triangular arbitragers. It is a characteristic of triangular arbitrage that it is cheap, fast, and almost unlimited in quantity. Thus, when all forward rates are in equilibrium in the sense of equation (3), any one forward rate in disequilibrium will be under very strong pressure from triangular arbitrage to return to its normal level. This operation involves a large number of currencies which by itself causes the depth of the triangular market. But, in addition to the large number of currencies, another factor tends to deepen the market even further. Whenever triangular arbitrage disequilibrates another market in the sense of equation (3), interest arbitrage funds will tend to restore equilibrium of the market in that sense. This tendency in turn provides further opportunity for triangular arbitrage pushing all rates into equilibrium in both the sense of equation (3) and of consistent forward rates.

From this reasoning, it may be concluded that an arbitrage schedule that includes triangular arbitrage is more elastic over its entire range than the AA schedule representing interest arbitragers alone. The inelastic part of the schedule is likely to occur at a greater quantity of forward commitments than if triangular arbitrage is disregarded.

Einzig reports [1] that London excels as an intermediary for many currencies such that a French businessman may find that he can buy a large sum of forward dollars for forward francs in London with smaller adverse effect on the rates than if he tried to sell the same amounts in Paris. This effect is due to what is essentially a triangular arbitrage operation. The operation in London is that dollars are bought with sterling and the French francs are sold for sterling.[2] This intermediary operation takes advantage of the depth of the £ markets and gives greater elasticity to the $/ffrs market.

The importance of including the action of triangular arbitragers in this analysis is that in contrast with the owner arbitragers and borrowers, their buying or selling activities do not necessarily involve spot currency and changes in reserves. Yet at the same time, they compete with and enter the market at the same end as owner arbitragers and borrowers. In so doing, they reduce the size of actual changes in reserves for any given change in forward rates.

(4) The Speculator-Trader Schedule

(4.a) Speculators

Speculation in forward exchanges involves a very simple operation. If a speculator expects the spot rate $E(X_o)$ 90 days from now to be different from the currently quoted 90 day forward rate (X_t), he enters a forward contract.

---

1. P. Einzig, A Dynamic Theory of Forward Exchange (London 1962), pp. 259-62.

2. If £·$ = 2.50
and £·F = 10.00
$·F must be (£·F)/(£·$) = 4.00
If $·F = 5.00, $ will be bought via the £ for 4.00 F and resold for 5.00 F.

If $X_t > E(X_o)$, he will offer to sell sterling forward because at the time when the contract matures, he expects to have to pay a smaller number of dollars for a pound sterling than he is receiving for it.[1] When $X_t < E(X_o)$, he will for analogous reasons offer to buy sterling forward. Theoretically, the speculator needs to own no funds for this operation, the fixed cost of the transaction is small. In practice, however, banks require the putting up of a margin.[2] If the money thus tied up earns no interest, speculation may carry an opportunity cost. The expected return from speculation E(R) depends on the size of $X_t - E(X_o)$. The greater this E(R) is, the greater the stock of forward exchange commitments entered by each speculator. In equilibrium the disutility from taking on the risk associated with contracting for the last unit of forward exchange is just equal to the gain from the expected rate of return.

Individual speculators do not usually expect any spot rate in the future to occur with absolute certainty. Instead, speculators usually hold expectations about a whole set of possible outcomes. Each of these possible outcomes carries a certain subjective probability of becoming reality and the expected rate ($E(X_o)$) can be considered to be the mean of that probability distribution. It holds in general that the greater the dispersion of possible outcomes around this mean, i.e., the greater the standard deviation (σ), the greater the possibility of large losses or gains. σ therefore can be regarded as an expression for risk. It is clear that the amount of forward contracts a speculator is willing to hold depends not only on the relation between $E(X_o)$ and X_t but also on σ. If the speculator is not a risk-lover, his willingness to hold speculative forward commitments is a decreasing function of σ. The parameters of this probability distribution are purely subjective. But if the individual demand and supply schedules are netted out, the resulting net market speculative schedule reflects the central tendency among the family of speculators. There is every reason to believe that it is much like that of individual speculators, and that it shifts in response to the same stimuli.

The accompanying Figure (3) shows these relationships geometrically. The horizontal axis, as in Figure (1), again measures the quantity of forward commitments while the vertical axis measures exchange rates. The schedule SS is associated with an expected future spot rate $E(X_o)$, and a risk estimate of σ_o. It goes through the point of zero forward commitments where the present forward rate X_t is equal to $E(X_o)$. The locus of points on the schedule indicates the relationship between actual forward rates and quantity of forward commitments resulting from the given $E(X_o)$ and σ_o. Thus, for instance, the forward rate X_t' results in forward sales of ON sterling by speculators in the expectation that they can be purchased spot at the lower $E(X_o)$ when the forward

---

1. J. Spraos, "The Theory of Forward Exchange and Recent Practice," Manchester School of Economic and Social Studies, XXI (May 1953), pp. 96-97. Tsiang, op. cit., pp. 86-92 for an algebraic treatment.

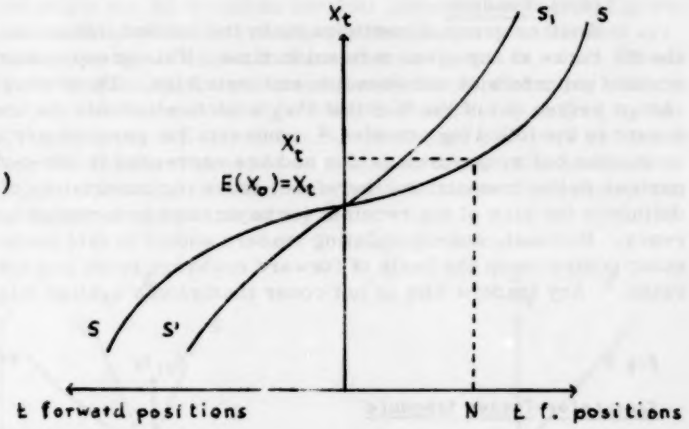
2. The Morgan Guaranty Trust Company of New York asked for a 10% margin in August 1961.

contract matures. At the extreme end, the SS curve becomes inelastic under the assumption of diminishing returns to risk taking. The schedule S'S' is associated with $E(X_o)$ and σ', where $\sigma_o < \sigma'$. The same risk estimate σ_o but a different $E(X_o)$ would shift the speculative SS schedule up or down parallel to itself.

Fig. 3

Speculators' Schedule

$S = f(E(X_o), \sigma_o)$
$S' = g(E(X_o), \sigma_1)$
$\sigma_o < \sigma_1$



(4.b) Dynamics of Speculation

The present analysis is essentially static. It tells what the amount of forward contracts entered by speculators is, given $E(X_o)$; σ_o and X_t. Dynamic analysis is required to explain how speculators reach a certain level of forward commitments as the SS schedule shows. For illustration, assume that a speculator sells forward ON sterling for delivery 90 days hence. Given X_t; $E(X_s)$; σ, he is in equilibrium. Ten days later he raises his estimate of $E(X_o)$ on the day of his maturing contract with σ unchanged. This means that his schedule is shifted upward in a parallel fashion and that he needs to lower his outstanding forward commitments to return to equilibrium. If a market for 80 day forward exchange exists, he buys sterling forward for delivery on the date of his original commitment. This act effectively reduces his net speculative position. It is possible that this can be achieved only by incurring a definite cost, i.e., the selling price under the 90 day contract may be lower than the buying price under the 80 day contract. But this cost may be smaller than the expected cost of keeping an open position after the events that changed the speculator's expectations. If all forward commitments can occur only in 90 day contracts, the speculator cannot reduce his open position as readily after a change in expectations as he was able to in the above example. He is then bound by a 90 day contract, a fact which enters his estimate of risk and thus determines the size of speculative engagements at any time. On the other hand, if the speculator's expectations about a specific date for which he owes sterling has changed such that he wants to lower this debt, there is a round-about way in which he can do it. He can simply buy sterling forward for delivery 90 days after his change

of expectations. If the two contract dates are not too far apart, and the change of estimates $E(X_o)$ on both days are closely correlated, the speculator's open position has been closed effectively. In the markets of today, forward contracts for odd days are readily available and the 30 day market is high developed. [1]

(4.c) Traders

Another group of participants in the market influences the position of the SS curve at any given moment in time. This group comprises the importers and exporters of merchandise and invisibles. Their need for forward exchange arises out of the fact that they wish to eliminate the exchange risk inherent in the following practice: [2] contracts for payment are normally drawn up months before payment is due and are expressed in one currency. One partner in the transaction, therefore, faces the uncertainty of not knowing definitely the size of his receipts (or payments) in terms of his domestic currency. Rational, non-speculating traders should in fact quote prices and enter contracts on the basis of forward exchange rates and not spot exchange rates. [3] Any traders who do not cover themselves against this exchange risk

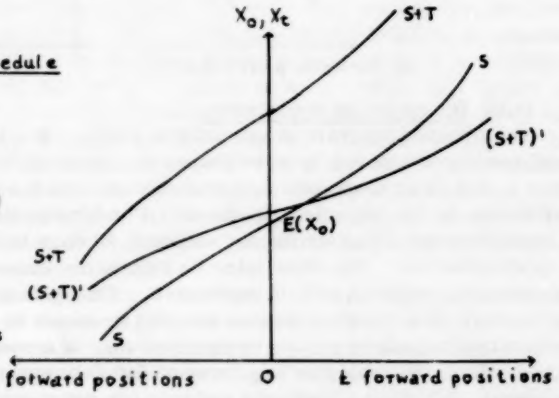
Fig. 4

Speculator-Trader Schedule

$S = f(E(X_o), \sigma_o)$
$S+T = g(E(X_o), \sigma_o, X, M)$
$X = \bar{X}$; $M = \bar{M}$

$(S+T)' = h(E(X_o), \sigma_o, X, M)$
$X = k(X_t)$
$M = j(X_t)$

X = exports
M = imports
$X_{t,o}$ = exchange rates



are considered, for analytical purposes, to be pure speculators because they expose themselves deliberately to exchange risks. Their open positions influence the location of the SS curve as described below.

In order to show how these commercial needs add to the demand and supply of forward currency, it is convenient to assume that all U.S. foreign

1. P. Einzig, A Dynamic Theory of Forward Exchange (London 1962), pp. 45-46.

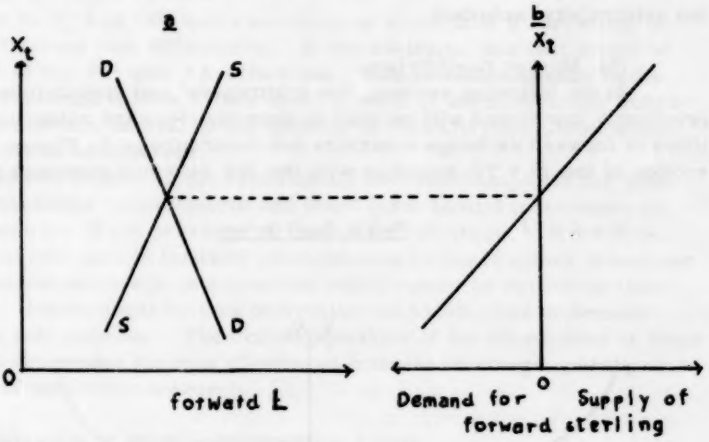
2. Note that this practice involves no spot currency or borrowing.

3. Lack of knowledge about the existence of forward facilities causes large proportions of commercial transactions to be uncovered. See Sohmen, op. cit., p. 25.

trade is billed in sterling and covered in the forward market. Two possibilities present themselves. One is that imports and exports are completely independent of the current forward rate. Since imports (M) give rise to forward purchases and exports (E) to forward sales, these can be "married," i.e., matched. But when $M > E$, an excess demand for forward sterling results. This excess demand shifts the SS schedule parallel upward as shown in Figure (4) to $(S+T)$.[1] The second possibility is that both the imports and exports are a function of the forward rates as shown in the rather inelastic D and S schedules in Figure (5a). The quantity of unmarried excess forward sterling thus is different in size and sign at the various X_t. The SS schedule is shifted to the more elastic $(S+T)'$ of Figure (4). At which X_t the D and S curves and therefore also SS and $(S+T)'$ intersect depends on domestic levels of income, demand and supply, elasticities of goods and services with respect to prices, etc. This topic goes beyond the scope of this study.[2]

Fig. 5

Demand and Supply of Forward Exchange from Traders



1. Reading, op. cit., pp. 307–08. His wording does not make clear whether or not he assumes the level of exports and imports to be independent of the forward rate.

2. For an analysis of how the demand and supply elasticities for forward and spot exchange rates are interrelated, see Sohmen, op. cit., pp. 72–75.

It is the assumption about the sensitivity of international trade to the forward rate which makes Kindleberger, op. cit., p. 335, ascertain that ". . . in the absence of speculation a rise in the interest rate will merely change the forward rate."

P. Einzig, A Dynamic Theory of Forward Exchange (London 1962), Chapter IX.

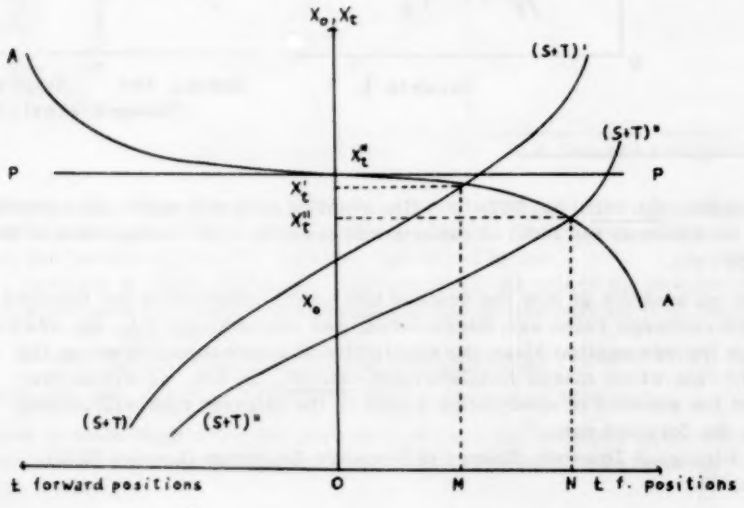
If the assumption that all traders cover themselves is dropped, the relevant consideration for the analysis is whether the proportion of all importers and exporters who fail to cover themselves is the same. The really relevant point is the difference between the demand and supply for forward currency by traders. It will be argued in part 5 of this section that when traders believe in a normal level of exchange rates, exporters are more likely to forego forward cover if the forward rate is above the normal spot rate, while importers react the same way when the forward rate is below the normal spot rate.

In summary, the S + T schedule indicates the relationship between the quantity of forward contracts outstanding given the expectations of all speculators $(E(X_o); \sigma_o)$ and the parameters underlying the import and export schedules. It is to be noted that this schedule indicates what quantities of forward sterling have been sold to the right of zero in Figure (4); the quantities purchased are measured on the left. This relationship is reversed in the case of the arbitragers' schedule.

(5) Market Equilibrium

In the following section, the arbitragers' and speculators' schedules previously introduced will be used to show how forward rates and the quantities of forward exchange contracts are determined. In Figure (6), the intersection of the (S + T)' schedule with the AA schedule produces a forward

Fig. 6 Market Equilibrium



rate of X_t' and a forward commitment of M. According to the assumptions underlying the (S + T) schedule, expectations and the balance of trade are such that at this equilibrium forward rate, speculators and traders are just willing to hold commitments for the delivery of forward sterling equal to M pounds. Interest arbitragers and triangular arbitragers, given the spot rate, the interest rate differential, and the other forward rates at their interest parity, hold at the forward rate X_t' commitments to accept delivery of sterling equal to M pounds.

A shift in the $(S + T)'$ schedule to $(S + T)''$ can be brought about by a change in expectations about the $E(X_0)$ or σ or both, or alternatively by an increase in covered trade surplus, or a combination of changes in these three parameters. The shift shown is sufficient, given the AA schedule, to cause the appearance of an X_t significantly different from X_t^* (i.e., causing an arbitrage margin greater than 1/2%). The forward contracts outstanding increase from OM to ON.

A change in X_t' and OM can also occur as a result of a narrowing or widening of the interest rate differential. In the diagram, this fact would be shown by a shift of the PP and AA schedules. Unless such a change in the interest rate differential causes a very specific shift of the speculative schedule, the new equilibrium as well as the quantity of forward commitments outstanding are likely to be different.

As was shown in part 2, the relationship between changes in the quantity of forward exchange commitments and short term capital movements is not one to one because of the presence of triangular arbitrage.[1] What proportion of a given increase in forward commitments is due to either triangular arbitrage or interest arbitrage is a question which cannot be settled on theoretical grounds. Institutional factors determine the elasticities of demand and supply from both sources. The critical question of the elasticities of these schedules which determine the size of covered interest arbitrage operations is thus a problem of empirical research.

6. Modification of Model—Intervention Points

(6.a) Intervention Points and Forward Rates

The theory of forward exchange presented so far assumes that spot rates are flexible without limits. Clearly, this assumption is not realistic. Under the IMF agreements, countries have obliged themselves to maintain spot exchange rates within 1% on either side of par.[2] During the fifties, par rates changed only rarely as countries showed new awareness of their responsibilities to prevent domestic instabilities and speculative runs through

1. J. Spraos, "The Theory of Forward Exchange and Recent Practice," Manchester School of Economic and Social Studies, XXI (May 1953), p. 91, warned that on these grounds, empirical judgment about the contribution of interest arbitrage in bringing the forward rate to parity is "very nearly impossible."

2. International Monetary Fund, Articles of Agreement, Article IV, Section 3.

responsible economic management. The longer these par values remain unchanged, the greater will be the confidence of the public in the maintenance of the accompanying intervention points. For the theory of forward exchange, this development has important implications which are based on the following fact: if an individual is very certain that the rate for sterling will not fall below 2.78 within the time period that concerns him, then when the rate has actually reached 2.78, it can only go up.[1] In the following section, the model of forward exchange determination will be modified to account for this fact.

(6.b) Arbitragers

The reasoning is carried out for the case where the spot rate is 2.78, i.e., at its lower intervention point (LIP) but is equally applicable to the case where it is at 2.82, its upper intervention point (UIP). When the spot rate is at 2.78, the parity forward rate can be above or below the spot rate. First examine the situation where the interest differential is in favor of London and therefore $X_o > X_t^*$. Given the assumption about the great confidence in the 2.78 floor, a movement of funds from New York to London without cover is encouraged. For by the time a London investment matures, the worst thing that can happen to an investor is that the spot rate has remained unchanged. Any appreciation of the rate represents a capital gain which accrues in addition to the higher interest earnings. Owner and borrower arbitragers stand to benefit in the same way.[2] But, does this mean that the demand from arbitragers will be zero and the AA schedule shrinks to nothing when X_o approaches 2.78?

The answer to this question is no. There are the triangular arbitragers who remain in the market. Furthermore, it must be realized that there is a significant difference between having great confidence in the maintenance of a floor to an exchange rate, and being certain about it.

As long as there is a non-zero probability of currency revaluation, the uncovered movement of funds is essentially speculation in spot rates. In the real world, two more considerations enter which tend to create a desire to cover. One is that a rate of 2.78 is reached rarely and only when the currency is in danger of requiring a basic readjustment. The second is that at small deviations from the limit, the risk from capital loss is greater than it may

---

1. O. Morgenstern, International Financial Transactions (Princeton University Press, 1960), pp. 166-69, for a discussion of how this fact enters the theory of the "international solidarity of money markets" under the gold standard. Reference is made there to studies showing that interest rate differentials between money markets normally did not exceed the greatest possible loss from exchange rate variation. This discussion disregards the existence of forward exchange facilities.

2. The shift of financing into the country threatened with depreciation of its currency is one of the ways in which the "lags" in the balance of payments is brought about.

appear at first sight. For instance, the depreciation of the exchange rate from 2.79 to 2.78 is only one cent. But when a one month investment is involved, such a small change represents a capital loss of 4.2% on an annual basis (i.e., 1/2.79 = .35% in one month = annual rate of 4.2%).[1]

Given these considerations, it seems reasonable to assume that demand and supply for forward exchange from interest arbitragers remains in spite of the fact that confidence in the exchange rate floor is great. This is so especially if it is correct, as has been argued above, that substantial covered interest arbitrage is carried out by large financial institutions. These mainly use borrowed funds and are known for their aversion to risk-taking and speculation. Borrowing arbitrage may retain its demand for forward cover in connection with the highly competitive, low profit margin trade in world staples. Therefore, the AA schedule remains essentially unchanged from its basic form derived in part 3. One important difference is, however, that the strongly inelastic range of the curve shifts to the right or left in response to changes in the spot rate. This is due to the fact that the closer the spot rate is to its lower limit, the more funds are moved without cover. This in effect reduces the quantity of funds available for covered arbitrage at any given margin to London as well as to New York.

Let us now turn to the case where $X_o = 2.78$, but the interest rate differential is in favor of New York and therefore $X_o < X_t^*$. Under these circumstances, it pays to move funds to London in riskless arbitrage only if the covered arbitrage margin favors it. Speculative movement of spot funds to London occurs when the combination of expected spot rate and the estimate of risk are such as to compensate for the certain loss in interest. The exchange risk of movements to New York remains and therefore forward cover is still required. The AA schedule under these assumptions retains its general shape, but again as above, its vertical section may move closer to zero when expectations favor speculation in spot funds which creates a smaller quantity of arbitrage funds at any given margin.

(6.c) Speculators and Traders

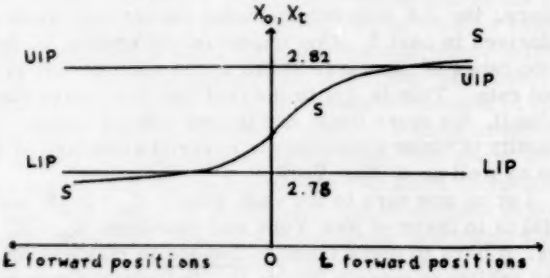
Speculators who believe confidently in the stability of par sterling at 2.80 should be willing to sell very large amounts of forward sterling when the forward rate is at the upper intervention point of 2.82. Analogously, they should be willing to buy forward sterling in large amounts when the rate is at its lower intervention point. For given their great confidence, the worst expected outcome would be that the spot rate on the future date is the same as the rate at which the future contract was made. At any spot rate between 2.78 and 2.82 a positive profit is made. For this reason, the SS schedule becomes almost perfectly elastic as the forward rate approaches the intervention points. On the other hand, the probability that a devaluation occurs is never zero. Entering an open position therefore carries a positive risk of loss which increases with the size of the open position. At some quantity of

1. Morgenstern, op. cit., p. 303.

forward commitments, the positive utility associated with the possibility of gain from buying or selling one more unit of currency is equal to the disutility of the additional risk arising out of this transaction. In order to coax the speculator into entering larger forward commitments, more substantial possible gain must be offered to him. Given expectations, this may occur only at a forward rate outside the intervention points. The section of the SS schedule between the ranges where it becomes almost perfectly elastic is also a bit more elastic than the one drawn under the assumption of perfect rate flexibility. It intersects the vertical axis where the expected spot rate is equal to the actual forward rate.

Fig. 7

Speculators' Schedule and Intervention Points



The demand for forward exchange by individuals engaged in international trade is also influenced by the existence of official intervention points and a widespread confidence in their maintenance.[1] A situation which typically confronts an importer is the following. The currently quoted forward rate is 2.82. If the importer believes with great confidence that the spot rate at the time of his payment will not be above 2.82, and perhaps lower, he is likely to gain an advantage by not purchasing the forward currency. The most unfavorable result is that he might have to pay 2.82 for the pound sterling when he settles his bill. But, there may also be the chance that he has to pay less than 2.82. Analogous reasoning applies to exporters and the lower intervention point.[2] As in the analysis of arbitragers and pure speculators, there is always some risk that a change in par value may occur. This risk causes some risk-averting traders to always cover themselves regardless of the odds regarding loss or gains from speculating on the exchanges. As compared with the group

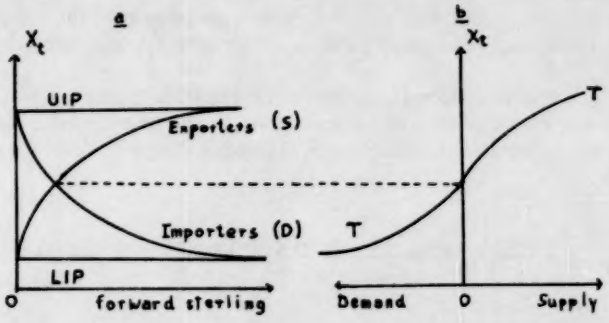
1. This fact has been considered by W. H. White in the Research Department Seminar of the International Monetary Fund. A resume of the material presented in this seminar was made available to me.

2. In a conversation, the head of the Foreign Exchange Department of the Chase Manhattan Bank of New York described this type of behavior by business men as typical. In a large number of cases, the decision to cover or not to cover is made on the basis of the Bank's recommendations.

of conservative financial arbitragers, however, the proportion of risk-avoiding traders is likely to be rather small. Thus, the demand for forward sterling by importers approaches zero when the forward rate is at the upper intervention points. For exporters, the function is a mirror image as shown in Figure (8a). The excess of demand over supply of forward exchange is therefore highly elastic as TT shows in Figure (8b).

Fig. 8

Demand and Supply of Forward Exchange from Traders



Underlying the functions in Figure (8a) could be the assumption that the level of exports or imports is completely independent of the forward rate and that instead it depends on the spot rate or some expected spot rate. Or alternatively, the assumption could be made that the level of exports and imports is determined by the forward rate all the time. The negative or positive excess of forward sterling demanded over that supplied would be greater at all rates than under the first assumption because in addition to the effect of speculation there is the normal effect that at high rates the supply for forward currency is greater than the demand as shown in Figure (5).

Combining again the SS and TT schedules results in the S + T schedule shown in Figure (9). The characteristic which distinguishes it from the one drawn under the assumption of perfectly flexible exchange rates is its greater elasticity over its entire range, including a long stretch of almost perfect elasticity near the intervention points. S + T indicates again what stock of forward sterling would be held by speculators and traders at any given forward rate. The schedule shifts in response to changes in $(E(X_o); \sigma)$ or the parameters underlying the traders demand and supply schedule.

(6.d) AA and S + T Schedules Combined

The model of forward exchange, modified by the assumption that intervention points and great confidence in their maintenance exist, is summarized in Figures 9 and 10. The schedules shown in these figures produce a relationship between forward exchange positions, forward and spot rates that differ from that of the basic model in the following ways:

1. When the spot rate is not close to the intervention point, the quantity of forward positions in equilibrium is likely to be much larger because of

the greater elasticity of the S + T schedule.

2. When the spot rate is very close to or at the intervention points, the effect of the greater elasticity of the S + T schedule is reduced by the shift of the inelastic part of the AA schedule toward zero. A given constellation of interest rates, spot and forward rates, and expectations about the future spot rate can result in a larger or smaller forward position than in the basic model.

3. When spot and interest rates are such that the parity forward rate is outside the intervention points, a covered arbitrage margin is likely to arise.

4. The relationship between forward positions and actual capital movements which is less than one to one in the basic model because of the triangular arbitrage is made more unpredictable by the uncovered movement of funds.

Market Equilibrium and Intervention Points

Fig. 9

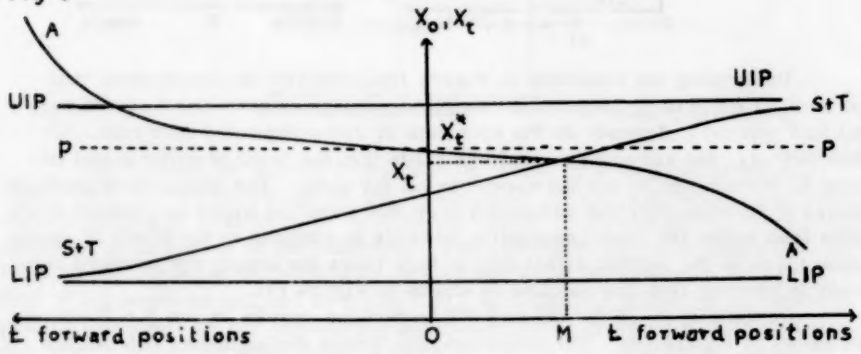
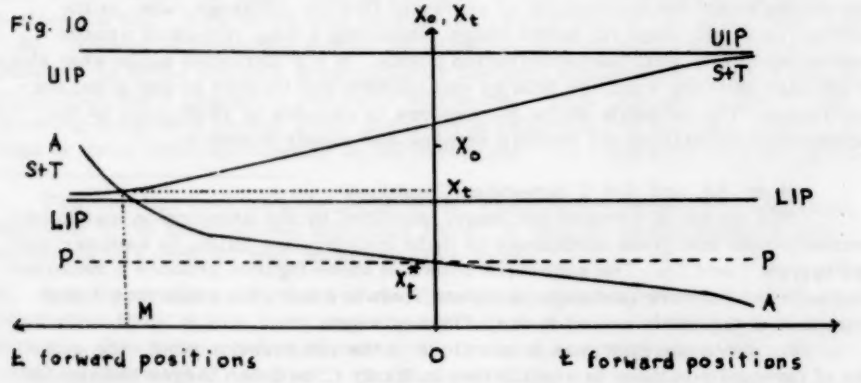


Fig. 10



(6.e) Proposed Widening of Intervention Points

The reasoning in this section makes it clear what would be the benefits of a widening of the range within which spot rates are allowed to fluctuate.[1] Whenever spot rates reach the margin, the potential capital loss from exchange rate change is increased which makes speculation less likely and covered arbitrage more likely. The number of times that the spot rate actually reaches the intervention points may be reduced. Furthermore, the margin by which short term interest rates can differ and yet not cause the parity forward rate to be outside the intervention points is increased for any given spot rate.

7. Maturity Structure of Forward Exchange Rates

The developed exchange markets of the world make it possible to enter contracts for the sale or purchase of foreign currency, delivery on the same day (cable), the following day (spot), or at any desired day in the future up to four years hence.[2] The best developed of these markets are the spot, 30 day, and 90 day markets. Daily quotations on these rates are readily available. Quotations for odd days and long periods are made by dealers and banks on special request.

What is the relationship between these exchange rates of different maturities at any given moment in time? The following analysis aimed at providing an answer to this question will deal only with the spot, 30 day and 90 day rates, and the relationship embodied in the concept of the implicit rate of interest, i.e. $((X_t - X_o)/X_o)(360/T)$.

The parity theory of forward exchange rates states that in equilibrium the forward rate is such that the implicit discount or premium is equal to the interest rate differential. This theory holds for 30 day as well as for 90 day securities and the corresponding forward rates. What is important is that the forward rates are determined by the differences in these interest rates. The question of the maturity structure of forward exchange rates is therefore very closely tied up with the question of the maturity structure of interest rates. It is customary to draw yield curves of securities as a monotonic, upward sloping schedule which indicates that the longer the time to maturity, the higher the yield.[3] If it is correct that yield curves have similar shapes in all countries

1. The widening of the gold points has been recommended by J. M. Keynes, Treatise on Money (London 1930).

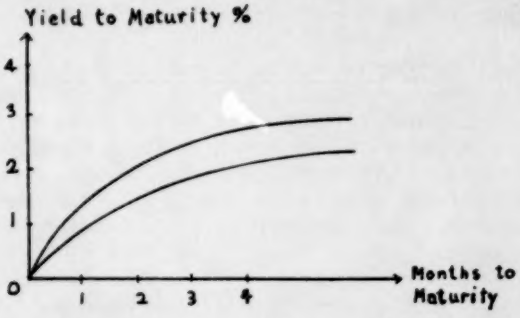
2. P. Einzig, A Dynamic Theory of Forward Exchange (London 1962), pp. 43–44.

3. P. Einzig, The Theory of Forward Exchange (London 1937), p. 228 ff. considers equality of implicit rates on all maturities as a sign of equilibrium. In Einzig's, A Dynamic Theory of Forward Exchange (London 1962), p. 249, he changes his position and says that ". . . since interest rates in the centers concerned vary according to the length of the credit the chances are that much more often than not interest parities too vary according to the length of the period." Einzig seems to refer here to the yield curve which in itself is not relevant. What counts is the distance between yield curves in the two centers.

and that they are shifted upward equidistantly over the relevant short term range when the level of interest rates goes up, then the differences between rates on securities with the same maturity should be approximately the same as shown in Figure (11). Given these assumptions, the implicit rate of return on 30 day and 90 day forward exchange should be equal. In the sense of the parity theory of forward exchange, they can be considered to be in equilibrium at that constellation. [1]

Fig. II

Maturity Structure



If the assumptions about the shapes and shifts of yield curves are changed, different implicit returns at different maturities result. The theory of forward exchange markets presented above provides further reasons why the implicit returns can be different even if the interest rate differentials are the same. If we can think of arbitrage schedules existing for each of the 30 day and 90 day markets because of the different needs of investors, then the theory implies that each forward rate is determined by the intersection of the arbitrage schedule with its corresponding speculator-trader schedule. It is possible that expectations about future spot rates are significantly different for 30 and 90 days and that therefore the equilibrium forward rate in each market produces a different arbitrage margin.

In a general way, it should be true that the speculative schedule for the 30 day market is more elastic than that for 90 days because expectations about closer dates are usually held with greater certainty (i.e., σ is smaller) than for distant dates. The conclusion from this is that the quantity of forward commitments in equilibrium is greater in the short market than in the long market, given the assumption that there are sufficient funds available for

---

1. See J. Conard, Introduction to the Theory of Interest (University of California Press 1959), pp. 287–346 for the theoretical foundations and empirical evidence justifying the drawing of yield curves in this fashion. Assume for convenience that all securities have to be held to maturity so that the following problem does not arise: the yield to some date preceding maturity exceeds the yield to maturity. See p. 325 for an explanation of how this phenomenon arises.

arbitrage in either market. For forward exchange policy, this means that when its goals are such that speculators increase the need for central bank commitments, the longer forward market should be used. The shorter market would be the better point of entry into the market when speculators' commitments are of the same type as those of the central bank.

If arbitrage is defined as the act of purchasing and selling the same commodity in two markets [1] in order to profit from price differentials in these markets, then arbitrage between forward rates of different maturities is not possible because the 30 and 90 day currencies are not the same commodity. Einzig describes the following actions as those of time arbitragers: ". . . sell three months forward and cover themselves by buying forward one month." [2] Einzig is clearly wrong to label these transactions "arbitrage." What is involved is double-barreled speculation. First, there is speculation that three months from a certain date, spot can be bought at a lower rate than that at which it is to be delivered. If this guess is wrong, it will not have helped that two months earlier there had been a contract to accept delivery of foreign currency under a forward contract for which payment could be made only by selling it spot. The profitability of this second speculative transaction stands independently from the 90 day sale.

Arbitrage involving forward sales alone on the other hand would be profitable and possible if three month contracts with one month remaining to maturity sold at a different price than one month contracts. But this is not likely to occur since the two contracts are equal in all important aspects and for this reason must be considered the same commodity.

III. Empirical Evidence and the Model

1. Introduction

Ideally it would be desirable to estimate statistically the most important parameters of the functions in the theoretical model developed in Section II. But unfortunately, lack of data does not permit such estimates through empirical research. The information required for such an ideal study would be data on spot and forward exchange rates, interest rates, quantity and type of forward commitments and expectations. Of these, it is data on the quantity and type of forward commitments which represent the big stumbling block. The absence of any statistics on this variable severely limits the scope and precision of empirical work on forward exchange.

Until such information is gathered and published, available data can be used to test the validity of theoretical considerations in a less precise and more general way. The following section describes the efforts of such general, empirical testing of the model presented in Section II. The first part of the section contains a general discussion of the data used. This is followed by an analysis of the results. A summary of the findings concludes the section.

---

1. K. Boulding, Economic Analysis (Harpers, Third Edition).

2. P. Einzig, A Dynamic Theory of Forward Exchange (London 1962), p. 253.

2. The Data

(2.a) General

In an effort to go beyond the ground covered by other empirical studies of forward exchange rate behavior,[1] data relating to more than two markets were obtained. These were rates quoted in New York, London, Paris, Ottawa, and Frankfurt. While such a study involved the handling of large numbers of observations, it was deemed profitable since the multilaterization of forward exchange markets as discussed in Sections II and IV entails important implications for forward exchange theory and policy.

The time period from July 1955 to July 1961 was chosen on the grounds that it contained periods of relative stability in exchange markets. It also contained strong speculative activity both before and after the introduction of convertibility of European currencies during the last week of 1958.

In all cases the raw exchange data were daily observations. These were averaged (simple mean) for weeks on the basis of which most computations for detailed analysis were carried out. Interest rates were weekly observations. The use of weekly rather than monthly or quarterly data was deemed necessary because the computation of triangular arbitrage margins would be rather meaningless for longer periods. Also in the computation of interest arbitrage margins weekly observations are likely to reveal developments which would be lost in the case of longer time periods.

(2.b) Spot, Forward, and Interest Rates

The spot and 90 day forward exchange rates were obtained from these respective markets: Frankfurt—DM·£, DM·F; New York—$·£, $·C; London—£·C, £·F.[2]

It was assumed that the most prevalent form of arbitrage involving the same exchange rate but two markets made it unnecessary to obtain both rates, for instance, $·£ as well as £·$. The validity of this assumption was tested and is described below. It was found to be a proper assumption.

For purposes of testing an hypothesis about time structure of forward exchange rates, 30 and 90 day rates for DM·£ were used.

In the case of Frankfurt, the exchange rate data are midpoints between the buying and selling rates quoted in the source. The London data are midpoints of the reported daily ranges. In New York, it was possible to obtain opening day buying rates, the use of which seemed most appropriate because

---

1. See part 3.a of this section for references to other empirical studies.

2. The meaning of the symbols is as follows: DM—Deutsche Mark; £—Pound Sterling; C—Canadian Dollar; $—U.S. Dollar; F—French Franc. DM·$ is the cost of one dollar in terms of DM.

The exchange rates for Frankfurt were obtained from the internal books of the Dresdener Bank, those for New York from the internal books of the Morgan Guaranty Trust Company. Exchange rate data for London came from Samuel Montagu's Review of Foreign Exchanges published bi-weekly in London.

they were quoted while the European markets were still open. In general, it is interesting to note that the difference between buying and selling rates in the case of Frankfurt forward quotations went up during times of crises and speculation, and became smaller during normal times. The range of London rates exhibited logically analogous behavior. No use was made of these variations in the computation of averages.

The weekly observations of interest rates [1] are subject to the same measurement difficulties as are obvious from the nature of the exchange rate data just listed. They do not reflect the difference between borrowing or lending rates, variations in price during the day or week, or special customer discounts. Such imperfections in data are typical in economics and they are handled here in the usual way. Some are taken into account in the statistical hypothesis and the interpretation of the results. Others are assumed to be insignificant.

No difficulty was encountered in finding meaningful interest rates on 3 months securities for New York, London, or Ottawa. For all three centers, rates on 3 months Treasury Bills were used. In addition, the rates on Bankers Prime Acceptances and Finance Paper were obtained for New York; those on Bankers Prime Acceptances, Prime Commercial Paper, and Day to Day Money were secured for London.

The Frankfurt money market is not very deep and well developed as compared with the London or New York markets. The Bundesbank, however, has in the past 7 or 8 years created a market for 3 month government securities which it sells at prices that are assumed here to reflect the opportunity cost of short term funds.

The problem of finding a meaningful interest rate for the underdeveloped Paris money market was more difficult. The Treasury Bill rate has been almost unchanged during the period under observation and presents a distorted opportunity cost for funds because these bills are used as legal reserve by French banks and special tax provisions make them especially attractive assets. Their yield was close to 3% throughout the period under consideration. The most meaningful market determined rate obtainable was that on day to day money secured by private securities.

(2.c) Information about Speculation

For the interpretation of some of the series computed, it is necessary to have some general information about the direction and vigor of speculative activity during this time. This information, together with some analysis, is contained in a few of the writings about forward exchange markets found in the References. The true flavor of the events, however, is best reflected in

1. The interest rates were obtained from the following sources: Monthly Reports of the Deutsche Budesbank, Frankfurt; Bank of Canada, Statistical Summary, Financial Supplements, Ottawa; The Economist, London; Federal Reserve Bulletin, New York; and an internal listing of the Bank of France, Paris.

contemporary articles in such publications as The Economist, Banker, Business Week, The New York Times, and The Wall Street Journal. Official publications such as the Federal Reserve Bulletin, Survey of Current Business, and Monthly Report of the Deutsche Bundesbank contain further references to the events under examination. Methodologically, it is not proper to search for indications of speculative activity only during those periods when the data hint at their existence. Care was therefore taken to examine the indices for references to unusual happenings in the exchange markets during times when the data suggested that there were none.

This more general information about speculation was buttressed by the collection of quarterly balance of payments estimates on private capital movements from Germany, Canada, the U. S., and England. These data are not included in this essay but can be found in the original version of the dissertation.

3. Computations and Interpretation

(3.a) Covered Interest Arbitrage

Based on the theoretical model of forward exchange rate determination developed in Section II, the following working hypothesis is to be examined: whenever expectations about future spot rates do not involve a revaluation of par, the intersection of the Arbitrage and Trader-Speculator schedules occur within the elastic range of the Arbitrage schedule. This means that empirically during normal times, we should find covered interest arbitrage margins very near to zero where this margin is defined as the difference between the implicit interest rate and the actual interest rate differential. (See Section II for a precise definition of these concepts.)

This basic approach to the study of forward exchange rate determination has been used in earlier studies of the subject.[1] Occasionally, official publications such as the Bank for International Settlements Annual Reports, Bank of Canada Statistical Bulletin, Federal Reserve Banks Monthly Reports publish series based on the same computations.

This study differs from earlier ones in several respects. It employs the concept of the precise arbitrage margin in the computations. The theoretical justification for this was presented in Section II; from a practical point of view, such a refinement was made feasible through the use of a high-speed computer. Furthermore, this study is different in that it is concerned with forward exchange rates quoted in more than one market. This has led to the compilation of some new statistics, presented in part (3.b) of this section.[2] The time period covered by the statistics is of special interest since

1. J. M. Keynes, op. cit.
   P. Einzig, The Theory of Forward Exchange (London 1937).
   B. Reading, op. cit.

2. O. Morgenstern, op. cit., has compiled triangular arbitrage margins for spot exchange rates for periods before the second World War.

it includes the transition from exchange restrictions to convertibility.

The reasons why empirical data cannot be expected to show absolute zero arbitrage margins even during normal times are as follows. There are measurement errors due to averaging of the daily and weekly variations, the use of buying or selling rates, disregard of quantity discounts, etc. in the case of both exchange rates and interest rates. Furthermore, transactions costs would prevent empirically computed arbitrage margins from ever becoming absolutely zero. In addition, there are the conceptual difficulties mentioned in Section II as to which interest differential to compare, i.e., that on Treasury Bills or that on Bankers Prime Acceptances?

Other conceptual difficulties result from reported switches from one type of securities in one market to another type of security in the other market (see Section II). Then there is the influence of triangular arbitrage involving forward rates alone, which could result in the appearance of covered arbitrage margins even when speculation between the two currencies is not going on. Lastly, we do not really know how elastic the arbitrage schedule discussed in Section II really is and at what range it is intersected by the speculator-trader schedule, even during normal times. But the fact that it is not perfectly elastic would explain some observed covered arbitrage margins even in the absence of these measurement errors and conceptual difficulties.

Given then in principle the likelihood for the discovery of covered arbitrage margins in empirical work, the question arises what size margin is to be considered reasonable and due to the above mentioned factors, and which is to be considered due to speculation. Earlier studies have chosen a margin of 1/2 per cent on either side of zero to be the cut-off point for the London-New York market. The choice of this 1/2 per cent was justified by reference to a survey of market participants made by Einzig in the thirties which had indicated this to be the critical point at which large numbers of arbitragers were drawn into the market because their transaction costs were covered.[1] In the current study, this transactions costs is to be treated as only one of many factors explaining empirical deviations from theoretically expected values.

Some such size as 1/2 per cent seems to fit the intuitive judgment about the general magnitude of measurement errors if only on the basis of its traditional use. But clearly it would not be the same in each market because the degree of competition among exchange dealers and bankers, the size of turnover in a currency, is different in each market and influences decisively the size of profit margins, i.e., margin between buying and selling exchange rates. The same reasoning applies to interest rates. How much such factors may have influenced the accuracy of the recorded observations and therefore the appropriateness of the 1/2 per cent margin has to be considered in the analysis of the computational results for each specific series.

1. P. Einzig, The Theory of Forward Exchange (London 1937).

The available data allowed the computation of six covered interest arbitrage margins involving Treasury Bill rates for the period 1955-61. They are presented on the accompanying charts, which also include representation of the movements of the implicit interest, and the actual interest differential. [1]

The space limitations imposed by this essay make it impossible to present a detailed analysis of the speculative episodes reflected in the accompanying charts. The interested reader is referred to the dissertation proper. May it suffice here to recall the following highlights of the events during the period under consideration.

The first episode was a bear attack on sterling connected with the armed intervention in Egypt during the Fall of 1956, after the nationalization of the Suez Canal Company in July 1956. The pressure on sterling was mainly due to capital leaving the country in fear of general war. The 1957 bear speculation on sterling, on the other hand, was prompted by the deterioration in Britain's balance of trade position and resultant expectations of devaluation. At the same time the French Franc had come under heavy bear-attack for similar reasons.

The year-end of 1958 marked the introduction of de jure convertibility of European currencies. The willingness to accept convertibility of these currencies has to a large degree been the result of the favorable liquidity position of the European countries. The largest source of this liquidity had been the US balance of payments deficits. This fact explains the source of the third major episode of instabilities during this period.

Fears of a dollar-devaluation after the 1960 election brought selling orders for dollar-balances. But during the latter half of 1960 sterling came under even stronger selling pressures than the dollar. Throughout this period of dollar and sterling weakness the Deutsche Mark attracted buyers in expectation of an upward valuation of the currency.

Inspection of the accompanying graphs reveals that these speculative episodes show up strongly in the accompanying graphs by large deviations of the arbitrage margins from zero.

(3.b) Triangular Arbitrage Margins

In equilibrium and disregarding transactions costs, the prices of dollars in terms of Deutsch Mark (DM·$) is equal to the product of the price of Sterling in terms of Deutsch Mark (DM·£) with the price of dollars in terms of Sterling (£·$). In symbols: (DM·$) = (DM·£)*(£·$) As defined in Section II, the algebraic difference between the two sides of the equation is the triangular arbitrage margin.

The available exchange rate quotations made it possible to compute triangular arbitrage margins for five combinations of rates, both spot and forward. These five combinations were: 1) (DM·$)-(DM·£)*(£·$);

1. Because the volume of the basic data is very large (i.e., over 12,000 weekly quotations) it is impractical to include a listing with the essay. The author is willing to lend a copy of the data to any party interested in them for further research.

2) (DM·$)-(DM·C)*(C·$); 3) (DM·£)-(DM·F)*(F·£); 4) (DM·£)-(DM·C)*(C·£); and 5) ($·£)-($·C)*(C·£). The graphical representation is contained in the next three pages.[1]

Theoretical considerations lead us to expect that these triangular arbitrage margins are always very close to zero. This is due to the fact that the market for currencies is very perfect, dealing in a homogeneous commodity with low transactions costs and well functioning channels of information. Thus, foreign exchange dealers are known to watch carefully any profit opportunities that might appear by the divergence of prices for the same currency in two different markets. Their simple bilateral arbitrage keeps prices equal between pairs of markets. Triangular arbitrage, similarly cheap and speedy, keeps prices consistent with each other.

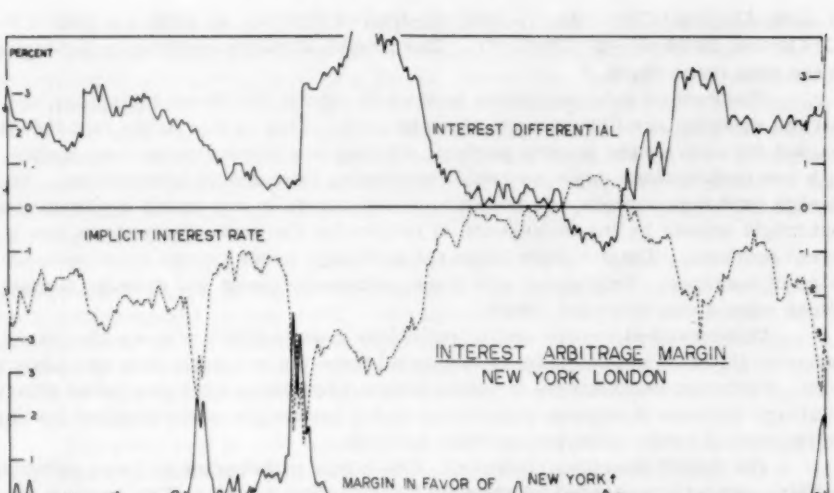
Measurement errors and transactions costs explain why we should not expect to find that profit margins computed from the available data are always zero. Exchange restrictions in effect before Christmas 1958 prevented effective arbitrage between European currencies and it seems plausible to allow for the appearance of profit margins on these grounds.

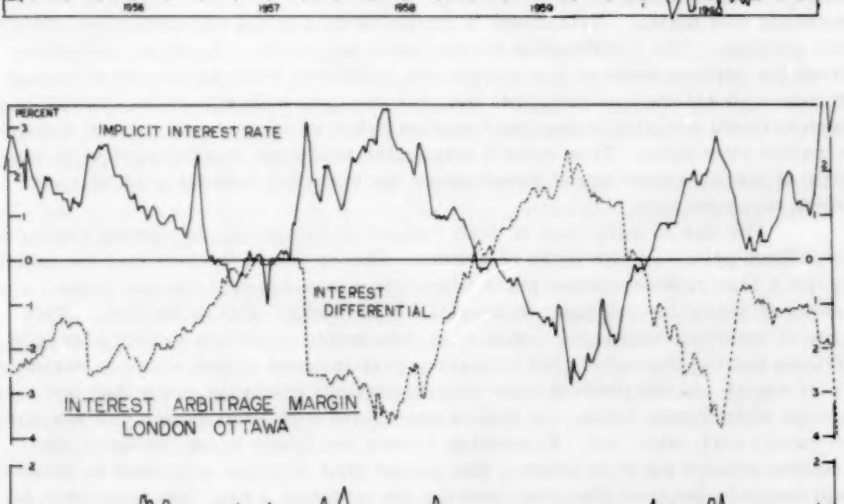
The question arises, however, which type of behavior in these arbitrage margins can be interpreted as showing the effective working of triangular arbitrage and which cannot. In order to find out whether triangular arbitrage was actually keeping forward exchange rates consistent, the following null hypothesis was tested. Deviations from the median triangular arbitrage margins are random. The justification for the basic test is the following: deviations from the median value of the margin are employed because the use of buying rather than selling (or midpoints between the two) exchange rates results in a bias between actual and observed market rates which was assumed to remain constant over time. Thus even if triangular arbitrage worked perfectly, this type of measurement error would cause the recorded margin to be of a constant, nonzero size.

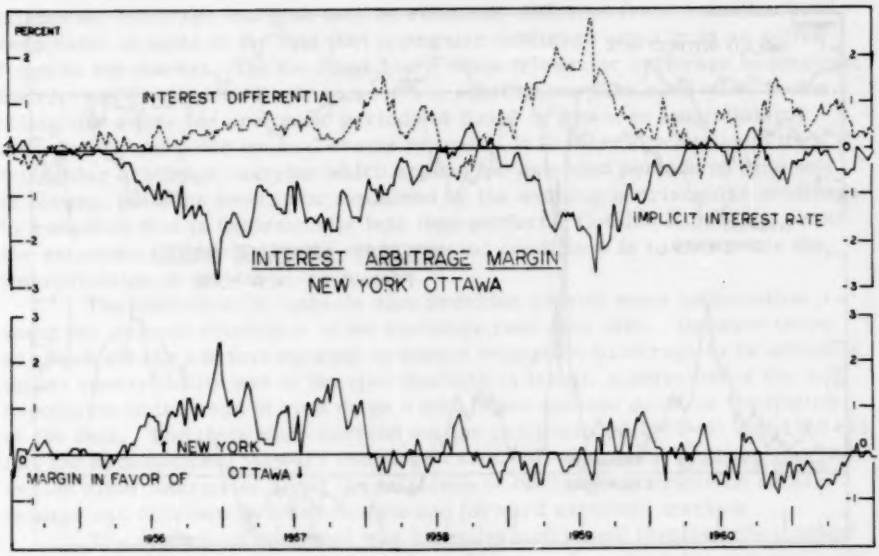
The use of midpoints of daily ranges in computing the weekly averages used here produces two types of errors. One is due to the fact that the midpoint is not a true representative price whenever a trend during the day caused a price above or below the midpoint to prevail for the longer part of the day. This type of measurement error tends to be minimized by taking weekly averages. Trends lasting throughout the averaging period cause a bias which is assumed to be small. In the present case of computations involving more than one such series of exchange rates, the biases associated with each series are assumed to cancel each other out. Remaining errors are likely to be randomly distributed around the true value. The second type of error is caused by quantity and special customer discounts making the midpoint a rate that may apply to only few transactions. This error is also assumed to cause random deviations of the observed rate from the true rate.

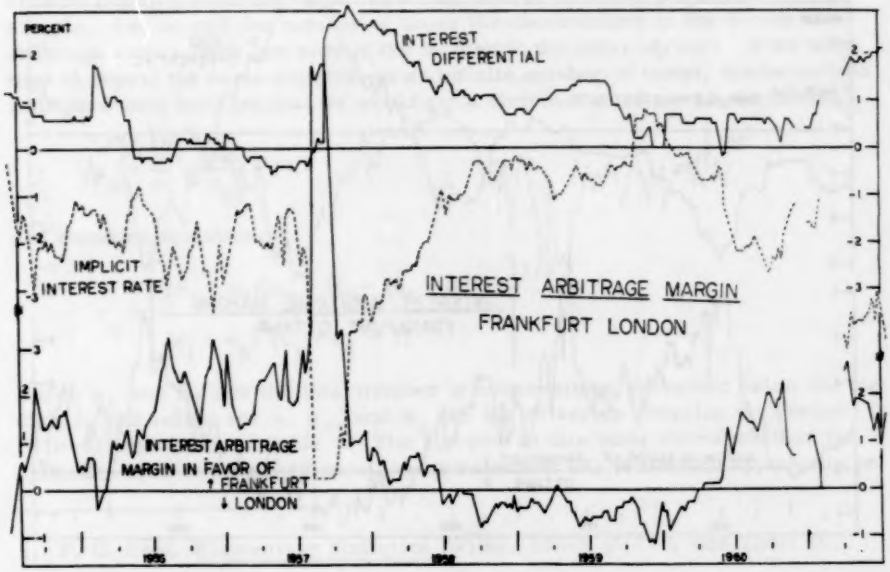
---

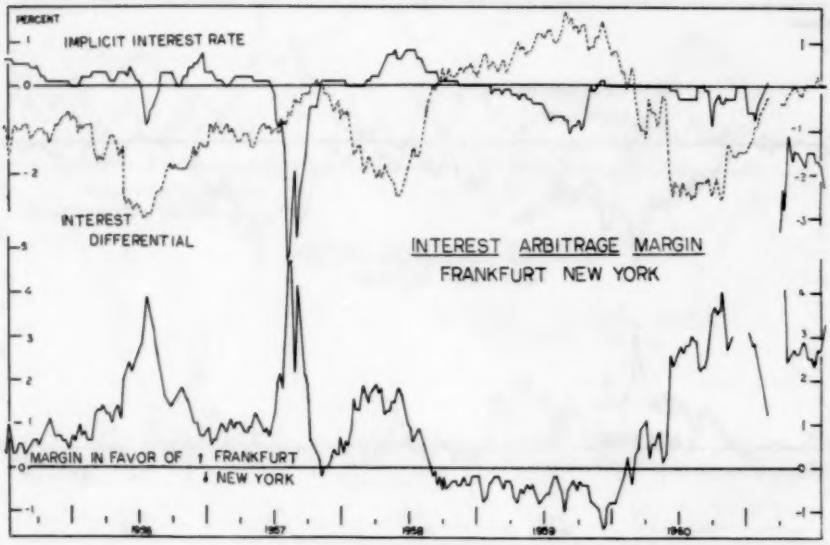
1. Again, the large volume of the data does not make it practical to include them in this dissertation. They can be obtained by writing to the author.

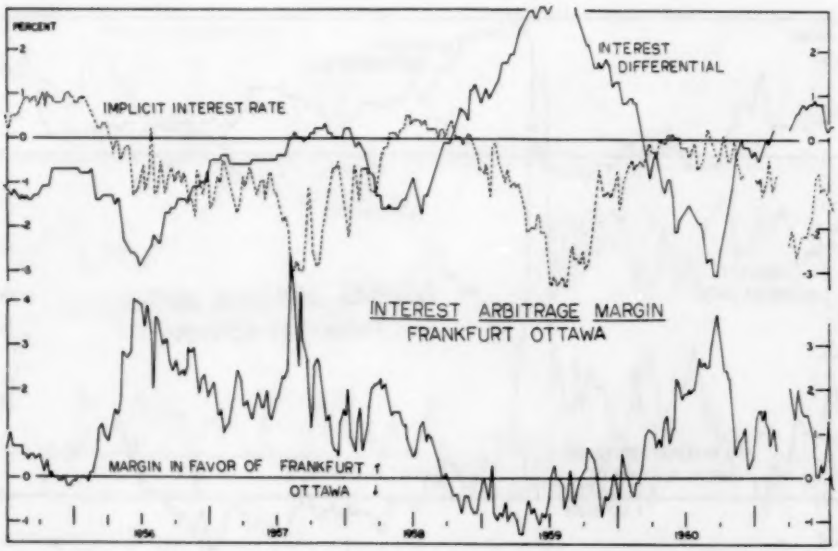

PERCENT
INTEREST DIFFERENTIAL
IMPLICIT INTEREST RATE
INTEREST ARBITRAGE MARGIN
NEW YORK LONDON
MARGIN IN FAVOR OF NEW YORK ↑
LONDON ↓
1956
1957
1958
1959
1960


PERCENT
IMPLICIT INTEREST RATE
INTEREST DIFFERENTIAL
INTEREST ARBITRAGE MARGIN
LONDON OTTAWA
LONDON ↑
MARGIN IN FAVOR OF OTTAWA ↓
1956
1957
1958
1959
1960


PERCENT
INTEREST DIFFERENTIAL
IMPLICIT INTEREST RATE
INTEREST ARBITRAGE MARGIN
NEW YORK OTTAWA
↑ NEW YORK
MARGIN IN FAVOR OF
OTTAWA ↓
1956
1957
1958
1959
1960


PERCENT
INTEREST DIFFERENTIAL
IMPLICIT INTEREST RATE
INTEREST ARBITRAGE MARGIN
FRANKFURT LONDON
INTEREST ARBITRAGE MARGIN IN FAVOR OF
↑ FRANKFURT
↓ LONDON
1956
1957
1958
1959
1960


PERCENT
IMPLICIT INTEREST RATE
INTEREST
DIFFERENTIAL
INTEREST ARBITRAGE MARGIN
FRANKFURT NEW YORK
MARGIN IN FAVOR OF ↑ FRANKFURT
↓ NEW YORK
1956
1957
1958
1959
1960


PERCENT
IMPLICIT INTEREST RATE
INTEREST
DIFFERENTIAL
INTEREST ARBITRAGE MARGIN
FRANKFURT OTTAWA
MARGIN IN FAVOR OF FRANKFURT ↑
OTTAWA ↓
1956
1957
1958
1959
1960

These measurement errors together explain why actually computed triangular arbitrage margins may be randomly different from a median non-zero value in spite of the fact that triangular arbitrage actually is an active force in the market. On the other hand, when triangular arbitrage is inactive, then forward rates are likely to be out of equilibrium with each other in the triangular sense for prolonged periods of time, or however long bilateral speculation and other unusual events prevail. It is therefore assumed that triangular arbitrage margins which appear for extended periods of time and in distinct patterns have to be explained by the working of triangular arbitrage in a manner that is theoretically less than perfect. Outside knowledge about the existence of speculation or other unusual conditions is to be used in the interpretation of the data.

The statistical hypothesis also provides us with some information about the general reliability of the exchange rate data used. Because there are such strong a priori reasons to expect triangular arbitrage to be effective (since convertibility and in the spot markets at least), a rejection of the null hypothesis in the case of spot rates would throw serious doubt on the quality of the data. The tests were carried out for two periods, 1955–61 and 1959–61 and for both spot and forward exchange rates. Comparisons of these four series allow inferences about the influence of exchange restrictions on arbitrage and differences between spot and forward exchange markets.

The statistical test used was non-parametric and involved the number and runs of deviations from the median. It is based on the following rationale: if a set of numbers were drawn randomly from a stable population, each successive drawing would be either equal to, above, or below the median of that sample. Let us call the number of times the observations in the series of drawings switch from one side of the median to the other u(runs). If we were then to repeat the same experiment an infinite number of times, mathematical considerations indicate that we would get a distribution of u with the mean:

$$\mu_u = \frac{2n_1 \cdot n_2}{n_1 + n_2} + 11$$

and standard deviation:

$$\sigma_u = \left[\frac{2n_1 \cdot n_2 (2n_1 \cdot n_2 - (n_1 + nn_2))}{(n_1 + n_2)^2 (n_1 + n_2 - 1)} \right]^{1/2}$$

where n_1 and n_2 are the total number of observations above and below the median.[1] The values for u, μ_u, and σ_u for the 10 series covering the period 1959–61 are shown in Table 1. The last row in this table shows whether the statistics lead to the acceptance (A) or rejection (R) of the null hypothesis on

1. P. G. Hoel, Elementary Statistics (Wiley, 1960), p. 170, and Table IX.

TABLE 1

Results of Tests for Randomness of Deviations from Triangular Arbitrage Median, 1959-61

Statistic	Frankfurt-New York London		Frankfurt-Ottawa New York		Frankfurt-London Paris		Frankfurt-Ottawa London		New York-Ottawa London	
	Spot	Forward	Spot	Forward	Spot	Forward	Spot	Forward	Spot	Forward
u	61	46	63	40	61	10	60	5	48	8
n_1	66	65	65	62	65	58	63	59	66	65
n_2	66	65	65	62	65	58	63	59	66	65
μ_u	67	66	66	63	66	59	64	60	67	66
σ_u	5.73	5.68	5.68	5.54	5.68	5.36	5.59	5.41	5.73	5.68
A or R [1]	A	R	A	R	A	R	A	R	R	R

1. (A) accept, (R) reject the null hypothesis.

the .05 level of significance involving a two tailed test.[1]

Inspection of the table reveals that the deviation from the median triangular arbitrage margin involving spot rates is indeed random with one exception. In the case of forward rates, however, factors other than random measurement errors seem to have influenced the pattern of profit margins.

Tests for randomness involving the entire period 1955-61 lead to a rejection of the null hypothesis in both spot and forward markets of all series with the exception of the one involving the spot rates Frankfurt-London-Paris. However, in all cases, a marked perfection of the markets in the theoretically expected sense is noticeable between the two periods. An examination of the accompanying graphs reveals this fact quite strikingly.

Frankfurt-New York-London

Peaks in spot and forward arbitrage opportunities appear during the 1955-56 speculative crisis as well as during the 1957 Suez crisis. The reduction of these margins occurs the week after the introduction of convertibility.

Frankfurt-New York-Ottawa

Casual inspection of the graphs can lead one to believe that both series show random fluctuations. The test for randomness, however, suggests that this is not the case. Only the spot rates after convertibility pass the test. With respect to the forward rates, a comparison of the two periods shows that the deviations from the expected median margin are relatively more random after convertibility than they were during the period as a whole.[2]

Frankfurt-London-Paris

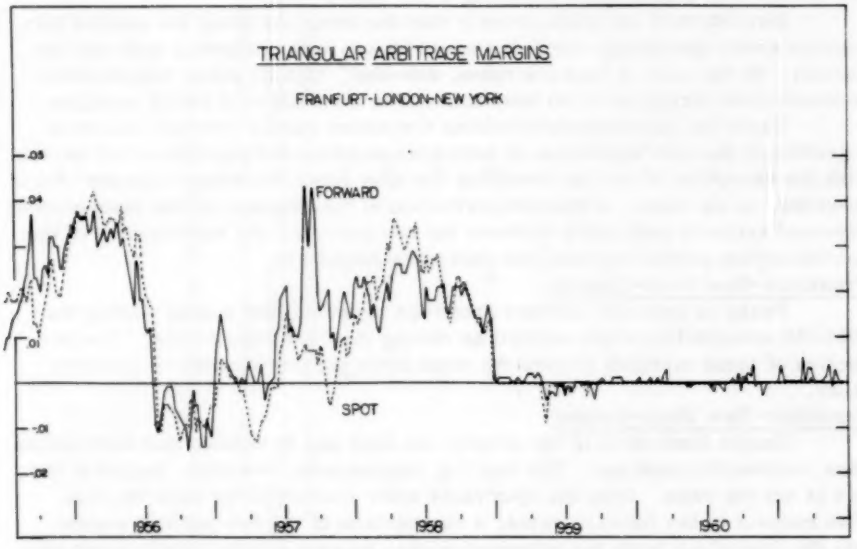
This series shows spot margins behaving in the theoretically expected manner throughout the period 1955-61. This may be interpreted to mean that while exchange restrictions can interfere with the smooth operation of the markets, the high turnover in the markets of all series with the exception of the one involving the spot rates Frankfurt-London-Paris. However, in all cases, a marked perfection of the markets in the theoretically expected sense is noticeable between the two periods. An examination of the accompanying graphs reveals this fact quite strikingly.

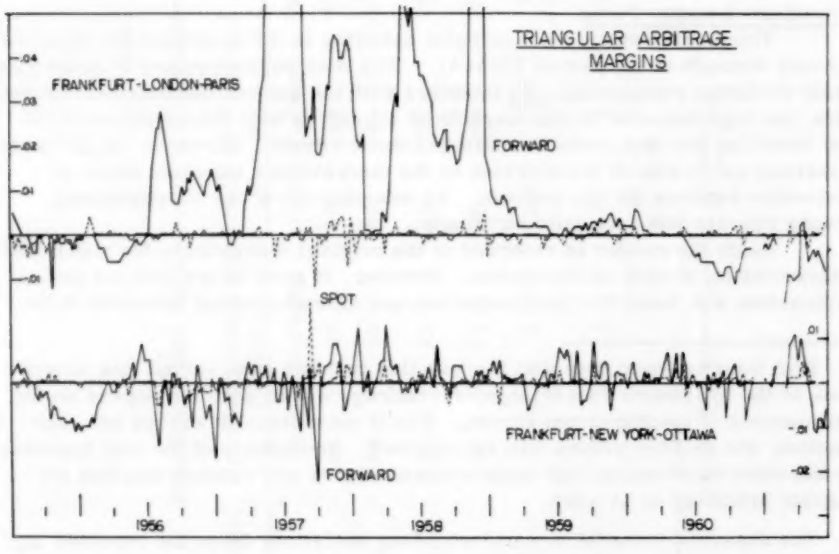
Again the reader is referred to the original dissertation for a detailed interpretation of each of the series. However, it must be pointed out that no explanation was found for the unexpected and almost cyclical behavior of the

---

1. It is interesting to note that the way this test has been set up, the acceptance of the hypothesis that triangular arbitrage works does not depend on the randomness of measurement errors. For if measurement errors are non-random, the null hypothesis will be rejected. Acceptance of the null hypothesis on the other hand means that measurement errors are random and that triangular arbitrage is at work.

2. The observed value of u was 7 standard deviations from the expected μ_u in the period 1955-61 while the difference was only 4 standard deviations for the period 1959-61.

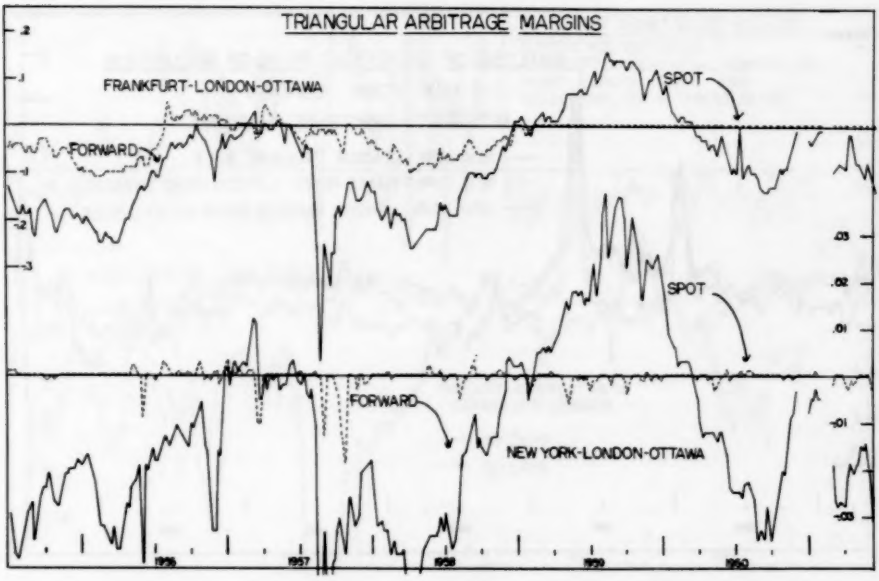

TRIANGULAR ARBITRAGE MARGINS
FRANFURT-LONDON-NEW YORK
.05
.04
.01
-.01
-.02
FORWARD
SPOT
1956
1957
1958
1959
1960


TRIANGULAR ARBITRAGE MARGINS
FRANKFURT-LONDON-PARIS
.04
.03
.02
.01
-.01
FORWARD
SPOT
FRANKFURT-NEW YORK-OTTAWA
.01
-.01
-.02
FORWARD
1956
1957
1958
1959
1960

triangular forward arbitrage margins in the series Frankfurt-London-Ottawa and New York-London-Ottawa.

In general, the data allow the following conclusions. Since the introduction of convertibility of the exchanges, the cross spot rates examined can be considered consistent. It is suggested that triangular arbitrage is the major force bringing about this consistency.

Fluctuations in arbitrage margins for forward exchange rates appear not to be random. In our analytical framework this means that triangular arbitrage is not the major force determining forward exchange rates. It is to be noted, however, that during the last two years under examination, arbitrage



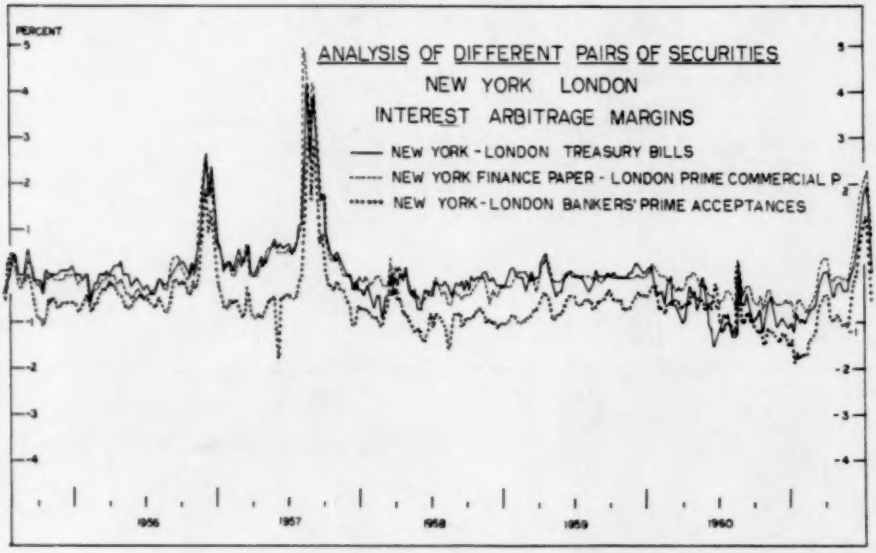
margins decreased in size and became more random as compared with the earlier period. This may mean that there is a tendency toward perfection of the forward exchange markets including more triangular arbitrage operations.

(3.c) Examination of Different Pairs of Securities

It has been possible to compare the interest arbitrage margins on three pairs of securities with equal maturity but different risk premia for the New York-London market. In the accompanying graph the arbitrage margins involving Treasury Bills, New York Finance Paper and London Prime Commercial Paper, and Prime Bankers Acceptances are shown. As discussed in Section II, if the securities are truly equivalent in their degree of riskiness above the risk attached to Treasury Bills in both markets, and if the rates are

all equally flexible, then the yield differential between them should be about the same. A different risk premium in one market on one security of the pairs should produce a differential that is unequal from that of the Treasury Bills by a constant margin. Since each observation of the three interest-difference variables is subtracted from the same implicit interest rate, the variations between them shown on the graph would also appear if they were just plotted without reference to forward rates.

The graph shows a great degree of correlation between the three series.[1] Two of the line intersect frequently. In terms of a sum of absolute deviations from zero the Finance Paper-Prime Commercial Paper differential (122.1) is



a superior measure of the working of forward parity theory than that of Treasury Bills (142.4). The difference between the first two series and the Prime Bankers' Acceptances (sum of absolute value of margins = 233.2) remains almost constant, as the graph shows very well.

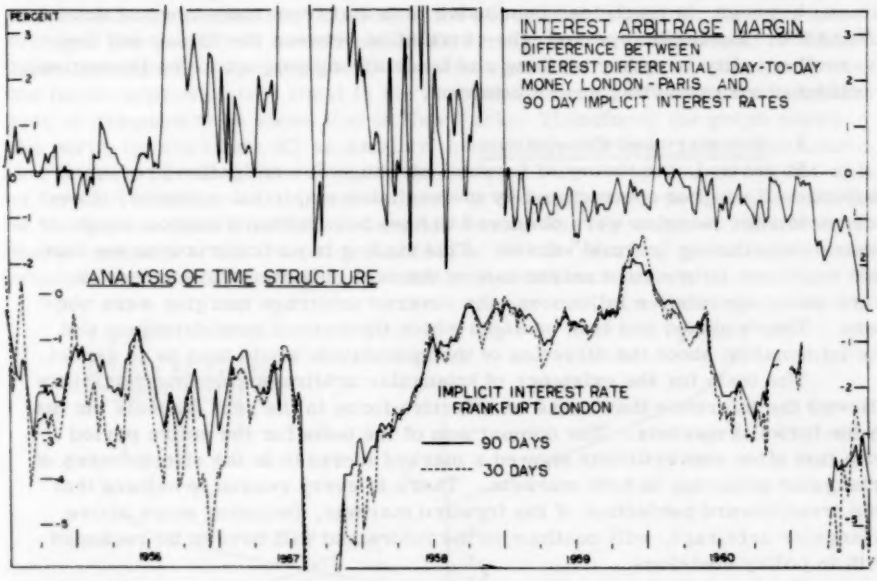
This seems to indicate that a different risk premium is attached to Bankers' Prime Acceptances in London than in New York. A more definitive judgment on this question requires further study of the institutional

1. P. Einzig, The Theory of Forward Exchange (London 1937), studied differences between discount rates, bank rates, rates on call money, and time money for the London-New York market 1920-36. His graphs exhibit a similar close correlation between these series.

arrangements which determine the risk-premia that markets attach to each type of security.

(3.d) Investigation of Time Structure

Since it was possible to get only interest rates on day-to-day money for the Paris market, it was decided to compute the covered arbitrage margin based on the difference of these rates with those on day-to-day money in London, using the three month forward exchange rates as the implicit interest rate. The theoretical justification for this procedure is, as was argued in Section II, that if the yield curves in two markets are similar in shape, and



shift equidistantly over the entire range when the level of rates changes, then the interest differentials on all securities of comparable riskiness will be approximately equal for all maturities. Thus the differential on day-to-day money should be about equal to that on 3 month bills for any two markets. Since three months forward currency is at parity with 3 month bills, it should be at parity with the spot rates.

The interest rates under examination exhibit wild day-to-day fluctuations both in London and in Paris, as compared with rates on longer maturities. The resulting covered arbitrage margins show equally wild gyrations during the years of exchange instability, from the middle of 1956 to the end of 1958, and early in 1961. Surprisingly however, during the remaining periods this margin remains rather stable, piercing a range of 1% only four times. In

1955-56 the remaining fluctuations, appearing almost random, are centered about zero. During 1959-60 they are centered about 1/2% in favor of Paris.

Another approach was used for the study of time structure by the comparison of the yield implicit on 30 and 90 day forward exchange. These rates were available for DM-Sterling and are shown on the accompanying graph. The premia or discounts over the spot rate were adjusted to an equivalent annual interest rate. The theoretical considerations of Section II lead us to expect approximately equal implicit rates for all maturities.

The graph reveals that during the periods of exchange instabilities the shorter rate fluctuates more widely than the longer rate but that they both follow approximately the same trend. The evidence from these speculative times, however, is much less conclusive than that from more normal times after 1958. During that period, the correlation between the values and their over-all equality is rather striking and lends strong support to the theoretical considerations predicting such a behavior.

4. Summary and Conclusions

In general, the theory of forward exchange determination presented in Section II was not contradicted by the available empirical evidence. Covered arbitrage margins were observed to have been within a narrow range around zero during "normal" times. This finding is as important as the fact that whenever information extraneous to the data indicated that currencies were under speculative influences, the covered arbitrage margins were nonzero. They widened and took on signs which theoretical considerations and the information about the direction of the speculation would lead us to expect.

The tests for the existence of triangular arbitrage since convertibility allowed the inference that it was an effective force in the spot markets but not in the forward markets. The comparison of the tests for the entire period with that after convertibility showed a marked increase in the effectiveness of triangular arbitrage in both markets. There is every reason to believe that this trend toward perfection of the forward markets, including more active triangular arbitrage, will continue in the future and will have to be reckoned with in policy decisions.

The investigation of different pairs of interest rates suggested fairly strongly that the theoretical consideration may be valid, i.e., it does not matter which pair of interest rates is examined as long as the securities are institutionally equivalent.

The findings with respect to day-to-day money of Paris and London are striking. Until a better explanation can be found, they support the theoretical hypothesis about the maturity structure of forward exchange rates. The same conclusion holds for the comparison of yields implicit in forward exchange rates of 30 and 90 day maturity.

IV. The Theory of Forward Exchange Policy

1. Introduction

This section makes use of the theoretical model developed in Section II

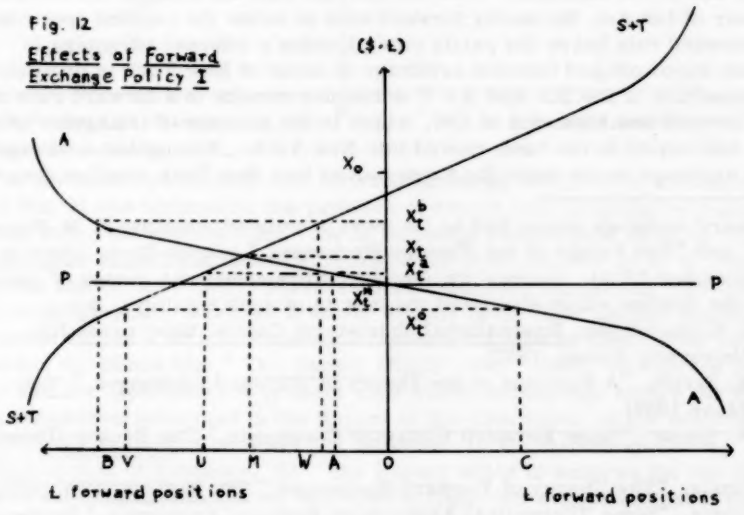
to analyze forward exchange policy, its potential and limitations. At first, the basic rationale for forward policy is demonstrated by the use of the diagrams analyzed in Section II. Then the institutional methods of intervention are discussed which is followed by an analysis of the influence of triangular arbitrage and how it can be overcome. This leads into an evaluation of the costs and benefits that can be derived from forward exchange policy during speculative attacks on currencies and during more normal times.

2. Basic Rationale for Forward Exchange Policy

In the accompanying Figure (12), the intersection of the speculator-trader and arbitrage schedules determines the forward rate X_t, the quantity of forward commitments, OM, and in the absence of triangular forward arbitrage, the quantity of investment in the foreign market, also equal to OM.[1] More specifically, the graph indicates that speculators expect a spot rate in the future approximately equal to the current spot rate. The London interest rate is assumed to be above that in New York. Therefore, the graph shows the parity forward rate X_t^* as smaller than the current spot rate. The London interest advantage attracts OM funds from New York covered by OM sale of forward Sterling which investors need to exchange the principal and interest of the London securities into dollars upon maturity. Speculators have bought these contracts at the low price planning to sell the Sterling at the higher price which is expected to prevail at the maturity of the contract.

Fig. 12

Effects of Forward Exchange Policy I



1. The model without intervention points is used here since it simplifies the exposition significantly. Intervention points are introduced in Part 8.

Forward exchange policy envisions influencing the quantity of foreign securities held by appropriate changes in the forward rate. Thus, for instance, starting out with the forward rate at X_t in Figure (12), changing the rate to X_t^a would cause a reduction in foreign net holdings of AM to OA, the increased forward rate X_t^b an increase of MB to OB. A forward rate of X_t^c would cause an accumulation of funds in New York of OC in spite of the fact that the original spot exchange rate, expectations, and an interest differential in favor of London were assumed to persist. Since investment in foreign securities means the exchange of spot currency against domestic currency, the rationale of forward intervention is obvious: a country's stock of foreign exchange can be increased or decreased at will, without requiring changes in the interest rate, thus leaving free this instrument for the achievement of domestic policy-objectives. [1]

3. Forward Exchange Policy and Speculation

While the basic principle of forward policy just presented holds in general for all constellations of exchange and interest rates, one case is sufficiently different to deserve special attention. This case is shown in Figure (13) and implies a bear speculation against sterling. This speculative sentiment is reflected in the S + T schedule which intersects the ordinate at a rate much below the current spot rate X_o at $E(X_o)$. At any current forward rate above $E(X_o)$ speculators are willing to sell sterling forward in quantities shown by the S + T schedule. With the interest rate differential assumed to be in favor of London, the parity forward rate is below the current spot rate. At any forward rate below the parity rate, London's interest advantage is more than wiped out and interest arbitrage in favor of New York is profitable. The intersection of the AA and S + T schedules results in a forward rate of X_t and forward commitments of ON, which in the absence of triangular arbitrage is also equal to the funds moved into New York. Triangular arbitrage in forward exchange would make the funds moved into New York smaller than ON.

1. Forward exchange policy had as its most prominent advocate J. M. Keynes, op. cit., and "The Future of the Foreign Exchanges," Lloyds Bank, Monthly Review (December 1935). Keynes' recommendations spawned a series of publications in the thirties which discussed the merits of such a policy. See:

C. P. Kindleberger, International Short-term Capital Movements (Columbia University Press, 1937)

W. W. Syrett, "A Revision of the Theory of Forward Exchanges," The Banker (June 1936).

W. W. Syrett, "Some Forward Exchange Comments," The Banker (December 1936).

P. Einzig, "The Theory of Forward Exchanges," The Banker (July 1936).

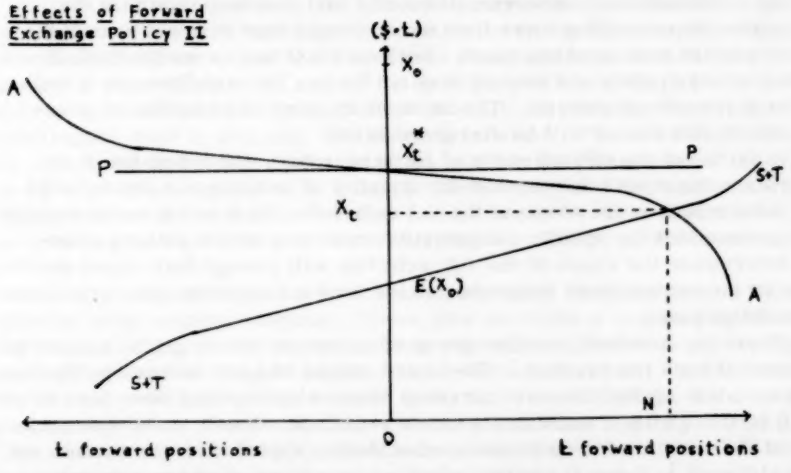
P. Einzig, "Some Theoretical Asspects of Forward Exchanges," Economic Journal, XLVI (September 1936)

P. Einzig, The Theory of Forward Exchange (London 1937). This publication also contains a history of forward exchange policies undertaken by governments.

As analyzed in greater detail in Section II, it is the action of speculators on the forward rate proper plus traders-turned-speculators (i.e., traders who don't cover themselves) who make up the sellers of forward sterling at a price which non-speculative arbitragers owning funds and borrowing find profitable for their operations. It is the loss of reserves caused by these arbitrage operations which can be prevented by forward exchange policy. But it it is true that the AA schedule is quite elastic over a wide range, substantial losses of reserves can be prevented effectively.[1]

Fig. 13

Effects of Forward Exchange Policy II



1. The literature concerned with forward exchange policy after the end of World War II was focused on the problem of how to curb speculative runs on sterling under the special conditions of exchange restrictions existing then. See the References for a detailed listing of the contributions by A. E. Jasay, J. Spraos and M. N. Trued. These authors concluded essentially that forward exchange policy would be a useful weapon in the fight against speculators. Opposition to this view came mainly from an article by an anonymous author, "Case for the Status Quo," The Banker (April 1958). Jasay and Spraos recently had the opportunity to present their views in hearings before the Radcliffe Committee published in the Report of the Committee on the Working of the Monetary System, Command 827 (London: H. M. Stationary Office, 1959), and Memoranda of Evidence, III. The Report failed to endorse the use of forward exchange policy during speculative times but suggested that it might be useful during times when speculative pressures were absent. See the Treasury memorandum, "The Forward Exchange Market—Policy," Memoranda of Evidence, I, pp. 121-22, for a point of view opposing the use of forward exchange policy against speculation.

4. Market Reactions to Intervention

A significant question concerns the reaction which the act of intervention itself can be expected to have on the behavior of buyers and sellers of forward exchange. Traders and speculators are most likely to change their expectations and the S + T schedule may shift. But it is hard to predict in which way the schedule will move. If traders and speculators believe that the intervention signals the government's willingness to maintain the status quo, they will react one way. If they believe the intervention to mean that the government admits there is a crisis and that the intervention is only a delaying action before the unavoidable revaluation, they will react in another way. Under any circumstances, however, the worst that can happen is that the government ends up holding more forward contracts than it would have if the negative reaction had not taken place. Such an event increases the cost of forward exchange policy and may or may not reduce its usefulness as a tool under the given circumstances. The issue of the cost and benefits of intervention which this raises will be discussed below.

In terms of the effectiveness of forward policy, the reactions of arbitragers are the most vital because the quantity of arbitrage at any forward rate is determined by the shape of the AA schedule. It does not seem reasonable to assume that the usually conservative investors whose pattern of behavior determines the shape of the AA schedule will change their asset preferences as the central bank intervenes in the forward markets and accumulates forward obligations.

There is, however, another group of investors who might be influenced by the central bank intervention. These are people who are induced to increase their speculation against the spot currency above what it would have been otherwise. It is this group of individuals whose reactions seem to make Auten conclude that the case for forward intervention during speculative attacks has not been established.[1] Auten fears that sizable accumulations of forward currency by the central bank may prompt very large speculation in spot currency which would embarrass the authorities and force devaluation, causing large losses on the forward contracts.

It is clear that any counter-speculation by a central bank, followed by a devaluation, would be disastrous. But as has been pointed out by all writers on the subject, forward exchange policy never is a substitute for devaluation. It is to be used only to combat speculation against an otherwise sound currency. It is the knowledge of this principle which signals to potential speculators that the central bank is determined to maintain par whenever forward policy is employed. Contrary to Auten's conclusion, therefore, the act of intervention itself is likely to be one of the most powerful tools with which confidence in a currency can be re-established.[2]

---

1. Auten, op. cit., p. 55.

2. See A. E. Jasay's statement on this point in the Principal Memoranda of Evidence, III (London: H. M. Stationary Office, 1960), p. 137.

5. Methods of Intervention

Discussion of the methods of intervention in forward exchange markets and what influence they may have on the determination of the rates has been neglected in the literature. Most writers limit themselves to a brief description of the two basic methods available. Yet Einzig implies that they may entail different effects and that therefore in any discussion of the merits of forward exchange policy it ought to be stated whether the use of one method of intervention or the other is contemplated.[1] It was therefore considered profitable to analyze in detail the effects of the two main methods of intervention available to central bankers.

In the first approach, the official agency enters the market as a buyer or supplier of forward exchange to raise or lower the rate respectively. What essentially happens is that the authorities stand by to buy or sell forward exchange at a specified rate, say X_t^a (see Figure 12). As the S + T schedule shows, speculators and traders want to buy OU sterling forward at that rate. Arbitragers want to sell only OA sterling forward. The authorities will therefore end up holding contracts which call for the future sale of AU sterling at X_t^a. In order to achieve an inflow of funds into New York against the interest differential calls for greater forward commitments because speculators as well as arbitragers want to sell forward sterling, say at X_t^c, in the total amount of OV + OC.

The second method of intervention available to authorities is that of offering swap accommodations. These take the form of standing by to sell (or buy) spot sterling at the current market rate while simultaneously buying (or selling) forward sterling at a stated rate. This forward rate is chosen to suit the objectives of policy with respect to the movement of funds, i.e., above parity, to effect a New York outflow of funds and below parity to effect an inflow of funds in our example. No takers of the swap offer will be had if the covered arbitrage margin offered by the swap-deal is smaller than that provided by the market determined forward rate. In order to see what happens when the swap-deal is actually profitable, assume that the policy aim is to increase the holding of London securities by New York investors and that swap accommodations are offered at X_t^b. Interest arbitragers will increase their demand for forward sterling at that rate to OB, which also represents the quantity of funds moved. The arbitrage schedule below X_t^b will become irrelevant as a locus along which the intersection with the S + T schedule determines the market forward rate. No arbitrager will be willing to buy sterling forward below X_t^b as long as the authorities are selling at X_t^b. This rate X_t^b thus becomes the market rate at which arbitragers become indifferent as to who sells the forward sterling they are buying. Speculators adjust to this rate and are able to sell OW forward. The authorities are left to hold WB, which is the same they would be holding had they sold forward outright.

The case where funds are to be moved to New York leads to the same amount of government involvement whether swap operations or outright sales

1. P. Einzig, A Dynamic Theory of Forward Exchange (London 1962), p. 389.

are used. The reasoning, however, is somewhat more involved. To see how it works, assume again the conditions underlying Figure 12. The authorities offer to buy spot sterling at X_o and sell forward at X_t^c. Since the covered yield on New York securities is now above that on London securities (i.e., $|X_t - X_t^*| < |X_t - X_t^c|$) OC forward sterling is sold by the authorities to cover the funds moved to New York. But now there are no more sellers of forward sterling in the market since all arbitragers have reversed their investments. As a result, speculators will offer to purchase forward sterling at a rate high enough so that covered London investments again become profitable. This offsets, at least in part, the original government action and can be prevented only by sales of forward sterling to the speculators-traders, either at X_t^c or at any rate above that, up to twice the distance of the X_t^c from X_t^*. Sales at rates other than X_t^c would mean different prices for forward sterling depending on their use and might not be easy to administer. A uniform price of X_t^c, however, carries with it government involvement equal to that of outright forward sales.

6. Triangular Arbitrage in Forward Exchange

Unitl now it was assumed that triangular arbitrage involving forward rates was not active. If this assumption is dropped, the AA schedule becomes more elastic and the relation between forward commitments and the actual flow of spot exchange is no longer one to one as was analyzed in Section II. For forward policy this has important implications which until now have been overlooked by other writers on the subject. In terms of the example used in Figure 12, the problem is this: if it was the object of forward policy to effect an inflow of funds into New York equal to OC, the forward rate had to be made X_t^c and the authorities' forward commitments were OC + OU. The presence of triangular arbitrage requires a forward rate smaller than X_t^c in order to achieve an equal inflow of OC funds. In turn this requires the authorities to hold more forward commitments from speculators as well as arbitragers. Potentially, this increase is quite large, its exact magnitude depending on the perfection and depth of the forward markets for all currencies.

Another aspect of the existence of triangular arbitrage is that if it worked well enough, a forward sterling rate pegged far enough away from parity would disequilibrate other forward markets, as described in Section II, even those not directly related to either dollar or sterling. While any intervention in the forward markets requires agreement between the two governments bilaterally involved, these multilateral effects may be serious enough to call for much more elaborate consultations before pegging by any bank.

Triangular arbitrage in forward exchange does not modify the conclusions reached under part (5) that in terms of forward commitments both intervention methods, direct sales and purchases of forward exchange, and swap operations are equivalent.[1] This may not be obvious because swap

1. This conclusion may be rather important. A high official of the foreign

offers require that anyone who wants to take advantage of the forward rate offered has to actually own spot funds and since triangular arbitragers do not own spot funds, they seem to be effectively excluded from taking advantage of this rate.

Triangular arbitragers can nevertheless increase the forward commitments held by the central bank above what they would be otherwise. To understand this consider the forward rate pegged at X_t^b (Figure 12). Triangular arbitragers would sell forward exchange to speculators and traders at that pegged rate. But this action reduces the amount of forward exchange which the speculators-traders would otherwise have bought from arbitragers. The central bank fills the gap of contracts offered by arbitragers at X_t^b and then ends up holding more contracts than before.

In this context it is interesting to note that the forward exchange intervention by the Federal Reserve Bank of New York early in 1961 (mentioned by Dillon in his speech cited in Section I) was confined to the forward DM. The Bundesbank had similarly limited its swap operations to the forward dollar. Since these interventions were largely aimed at reducing margins opened by speculators, they should not have led to large-scale triangular arbitrage operations.

7. Pegging More than One Currency

Until now the analysis has proceeded under the assumption that the government pegging operations were confined to one currency. This is obviously not necessary and the multilateralization of intervention removes some of the complications brought about by triangular arbitrage.

What is involved is this, the equilibrium condition for triangular arbitrage is: ($\cdot$£) = ($\cdot$C)*(C$\cdot$£). If now the ($\cdot$£) rate is pegged at a value which makes the equality an inequality, then triangular arbitrage tends to exert pressure on both the ($\cdot$C) and (C$\cdot$£) rates which produces the undesirable results discussed under part (6). This occurrence can be prevented, however, if the ($\cdot$C) rate is changed by the same percentage as the ($\cdot$£) rate, thus preserving the equality. If all dollar forward rates are in this way pegged an equal percentage distance from their interest parity, multilateral arbitrage in forward rates is not profitable. The quantity of forward commitments entered by the intervening government agency will again reflect more nearly the true amount of spot funds moved.

The elimination of the bad effects of triangular arbitrage, achieved by the pegging of all rates carries a cost. First, there is the administrative

---

exchange department of the Federal Reserve Bank of New York suggested to me in a private conversation that central banks could use swap operations rather than outright sales as one method of dealing with the problems arising out of triangular arbitrage. He thought that this would lead to two sets of forward exchange rates in the market. One of them would apply to owner-arbitragers and the other to the free market. The latter rate would presumably be consistent with other forward cross rates.

difficulty of dealing with many currencies. Second, there is the more complicated round of negotiations between the much larger number of countries thus involved before intervention can take place. The reader is reminded of the chaos that may result if forward policy is used widely and the central banks in two countries peg the rates at different levels. Alert private operators could inflict staggering losses on these banks.

It must furthermore be emphasized that to be really effective, forward exchange policy has to be used promptly and in fine gradations, depending on local developments. Lengthy multilateral negotiations on an ad hoc basis are likely to be slow, cumbersome, and inadequate to guarantee the effective working of the policy. Therefore, unless international agreements can be reached and channels of communication institutionalized in advance of actual need, the usefulness of forward exchange policy is seriously reduced.

8. The Costs vs. the Benefits of Intervention

Any discussion of the costs and benefits of government activity must distinguish between purely financial costs which are essentially transfer payments, and real costs and benefits which involve the use or creation of actual resources.

The financial costs of government intervention in the forward markets arises out of the difference between the price at which the currency is sold forward and at which it has to be purchased spot when the contract matures. This cost may in effect be zero if the central bank holds spot exchange in its portfolio which it acquired through intervention in the market when the foreign currency was near its lower intervention point. But if no such exchange reserves exist, the probability is great that under normal circumstances (i.e., no speculation in anticipation of devaluation), there is a financial loss involved. The size of this loss depends on the future spot rate and the size of the commitments.

In order to give greater precision to the variables which determine the cost of forward intervention, consider the following. Assume that the elasticities of the S + T and AA schedules are as shown in Figure (12) with covered arbitrage funds of OM having moved. Let the policy objective be to increase the amount of funds to OB. This requires pegging the 90 day forward rate at X_t^b. It is expected that the need for the increased foreign funds of MB will last for 6 months and that none of the parameters underlying the schedules in Figure 12 will change during that time. The cost of keeping these funds for 6 months will be:

$$C = (X_{90}^b - E_{90}(X_o)) \cdot (V + D)$$

$$+ (X_{90}^b - E_{180}(X_o)) \cdot (V + D)$$

In terms of the example, this means that the cost will be equal to the following: 90 day forward rate pegged at X_t^b minus spot rate 90 days hence multiplied by the quantity of funds to be attracted (V = BM) plus the quantity of

forward contracts to be taken over from the reduced holdings of speculators (D = MW). To this must be added the cost of repeating the process once more at the end of 90 days.[1]

From the above analysis, it should be clear that the magnitudes or parameters of none of these variables are actually known to the central banker. All he is reasonably sure of is the magnitude of the required capital inflow, V. The elasticities of the two schedules are unknown, as is the influence of triangular arbitrage, and whether and how the schedules shift in response to the pegging operation. This makes it impossible to know precisely at what rate the pegging has to occur, or how large D will be. The actual spot rate 90 and 180 days hence can also only be estimated. And, of course, the above assumptions about the constancy of the relevant parameters of the functions are unrealistic.

The analysis of forward exchange policy presented in parts 2-4 of this Section allows, however, some rough comparisons about the financial cost of forward exchange policy under various circumstances. Quite striking is the jump in cost that can be expected from efforts to make arbitrage funds flow against the existing interest differential (see Figure 12), as compared with mere attempts to attract funds into the high interest center. It is noteworthy to point out the especially crucial role of the elasticity of the S + T schedule in speculative situations.

The fact that the financial cost of forward exchange policy cannot be ascertained in advance, however, should not be a deciding factor in the choice of whether or not to intervene in the forward markets. Instead, the expected gains in real terms from such a policy should be decisive. These benefits, during times when no speculative pressures exist, arise out of the fact that interest rates can be kept lower than would otherwise have been the case. This carries not only monetary but real benefits. Lower interest cost to the government treasury on servicing the public debt represents a financial saving which could be used to offset the losses from the exchange operations in an accounting sense. If any part of the debt is held by foreigners there may be a saving in real terms because of the lower interest rates. On the other hand, part of this saving would tend to be offset by increased holdings of the debt by foreigners as a result of the interest arbitrage, and possible earnings from speculation by foreigners.

The real benefits from the lower interest rate come from the opportunity to stimulate investment, growth, and maintain full employment, the achievement of which presumably is the goal of the low-interest policy and necessitates forward exchange policy in the first place (see Section I). This full employment and increased output yields returns which should carry the heaviest weight when deciding on the use of forward exchange policy.

During times of speculative attacks on currencies the potential gains from the use of forward exchange policy are also rather large in real terms.

---

1. This approach to the presentation of the costs of forward exhange policy has been suggested to me by Professor Michael Lovell of Yale University.

They arise out of the possibility that intervention in the forward markets can prevent devaluations that might otherwise be forced by speculators. And since devaluations and currency instabilities are detrimental to international trade and the benefits derived from it, the avoidance of frequent currency adjustments would yield large real benefits.

Very sizable financial losses by the central bank, however, cause redistributions of income which might be undesirable from a social point of view. And undoubtedly there exists a point where loss in welfare accompanying these redistribution effects is greater than the gain in welfare derived from the policy itself. While, as a matter of empirical judgment, there is little danger that such a point will ever be reached, the possibility of its occurrence must be taken into consideration by those deciding on the use of forward exchange policy.

No discussion of the cost and benefits of forward exchange policy is complete without reference to the individuals who are normally responsible for its execution. They are central bankers, known for their conservative attitudes. These attitudes may lead them to attach quite special valuations to some of the effects, costs, and benefits of forward exchange policy. There is the basic principle of interfering with the free forces of the market which, however, has lost much of its strength as an argument since central banks throughout the world have taken on the responsibility for maintaining stable economic conditions. Then there is the obviously speculative air of the operations, allowing only rough and unreliable estimates of the cost and some of the effects of forward exchange policy. Other central bank operations suffer from similar defects but few of them are associated with financial losses with as much certainty as forward exchange policy. These losses appear as real numbers in the banks' balance sheets while the potential gains for the economy are more anonymous. Also it is the bankers who have to see that the possibly tedious international negotiations bring workable agreements.

It is therefore quite understandable that in the opinion of some central bankers the costs of forward exchange policy may not be worth the benefits. It can only be hoped that in time rational considerations and the availability of reliable estimates on costs, etc. will succeed in convincing bankers of the potentially large benefits that can be derived from an imaginative use of forward exchange policy.

9. Intervention Points

Until now the analysis assumed that spot exchange rates were free to fluctuate without limit. The removal of this assumption has two main effects, as analyzed in Section II. One is that the AA schedule becomes inelastic at smaller quantities of forward commitments when the spot rate is near an intervention point. The second is that the S + T schedule is almost perfectly elastic around the intervention points.

Whenever the forward rate to be pegged lies inside the intervention points or is not too close to them, these modifications of the basic model carry only minor implications for forward policy. The quantities of forward contracts

held by the central bank are likely to increase as a result of the more elastic S + T schedule and there may be a limit on the amount of covered arbitrage funds because of the inelastic part of the AA schedule. Whether and how much these factors reduce the usefulness of forward exchange policy by increasing its cost out of proportion to the benefits is a matter of empirical judgment which only actual working in the market permits. It is the author's opinion that the AA schedule is probably elastic over a wide range in most markets and that the S + T schedule is not so elastic as to cause serious harm, especially when the policy does not try to move funds against the interest differential.

During speculative times when the par value of a currency is under suspicion the intervention points lose much of their meaning because a change in par brings a change in these points. The S + T schedule therefore ceases to be nearly perfectly elastic around these rates.

During normal times, however, when there is a strong belief in the permanence of the par exchange rate and the accompanying intervention points, the usefulness of forward exchange policy may be seriously impaired if the forward rate has to be pegged near or outside of the intervention points. The result of such pegging at the critical rates will be that very large quantities of forward contracts are likely to be offered to the central bank. What this means in terms of financial losses to the intervening government agency has been discussed above.

The implications of this analysis for the general usefulness of forward exchange policy are rather clear. There undoubtedly exists a point where the effects of large financial losses on the distribution of income are socially more undesirable than the unemployment which would exist if the policy were not used. It is an empirical guess that such a point is likely to be reached when the forward rate is pegged in the range where the S + T schedule becomes almost perfectly elastic.

But all this does not mean that forward exchange policy becomes useless under these circumstances. Several methods of dealing with the problem are available. The most widely recommended is that intervention points be widened. This probably represents the most promising long-run approach to the alleviation of the problem in that it would tend to reduce the number of times when the pegged forward rate is near these points. Its main drawback is that it requires a basic change in the IMF agreements. This may not be easy to negotiate.

Another way to alleviate the problem lies in entering the market in the shorter maturities. For instance, to overcome an interest differential between two centers of 4%, a spot exchange rate at 10.00 would require a forward rate of 9.90 (or 10.10) in the 90 day market but a forward rate of only 9.967 (or 10.033) in the 30 day market. If the official spot intervention point were 9.95 (and 10.05) central bank pegging in the 90 day market would cause difficulties which pegging in the 30 day market could avoid.

Another promising method of dealing with the difficulties arising out of the existence of intervention points stems from the following fact. The

covered arbitrage margin, the size of which determines the magnitude of the capital flows, is a function not only of the forward exchange rate but also of the spot rate, and the domestic and foreign interest rates. Insofar as central banks are willing to influence the spot rate through purchases or sales of foreign exchange, they can achieve the desired arbitrage margin with a forward rate pegged away from the critical region. Often it may be convenient to adjust the interest rate somewhat in the direction which would restore the usefulness of forward exchange policy. The interest rate adjustment required for external stability is still likely to be smaller than it would have to be in the absence of forward exchange policy.

To summarize, intervention points and confidence in their maintenance can be considered complicating factors in the administration of forward exchange policy under certain circumstances, but they do not substantially reduce its over-all usefulness if accompanied by the appropriate and complementary policies discussed above.

10. Conclusions

The above analysis represents a rather clear cut case in favor of active forward policy during times of speculative pressures on a currency, when the exchange rate is basically in equilibrium and requires no revaluation. The use of forward exchange policy, on the other hand, is not to be recommended in speculative situations where a fundamental balance of payments disequilibrium exists. It can bring some temporary relief from speculative losses in reserves, but this period of reprieve is likely to be very expensive once the unavoidable depreciation takes place.

The economic case for forward exchange policy during periods when no speculative pressures exist is also rather clear cut. It is most strongly so when there is a need for reinforcement of flows in the direction of an existing interest advantage, but less so when there is a need for the reversal of a basic interest advantage because of the larger costs involved in the second case. In general, the economic case will become increasingly stronger as institutional developments tie the money markets of the world closer together and make the AA schedule more elastic, and as the maintenance of par exchange rates and intervention points causes a greater readiness of traders to behave in such a way that the S & T schedule also becomes more elastic. These two effects together mean that given interest differentials will in the future cause increasingly large shifts of arbitrage funds which may eventually put pressures on international reserves of countries and call for the adjustment of interest rates or the use of forward exchange policy. It is possible that the point has already been reached in these developments where the economic advantages gained from forward exchange policy are great enough to outweigh the disadvantages of administrative and financial costs, required complementary policies, and the need for international agreements.

A word of caution is required in conclusion. Forward exchange policy is not a panacea to the problems of speculative runs on currencies, or the shift of vast sums of short-term funds from one country to the other. As is

implicit in the model presented, speculation may anticipate correctly the needed change in the exchange rate. Effective pegging may often have to be at rates very much outside the spot intervention points. Large proportions of the foot-loose short-term money may never look for forward cover. This might make these funds insensitive to forward policy and throw the burden of adjustment back on interest rates and a wise mixture of fiscal policy and other methods. But as long as forward policy is useful at all, and this study has shown that it can be under a wide range of conditions, then it should be included in the kit of tools available to people responsible for stable national and international monetary affairs.

Summary: The study examines wage rate and hourly earnings changes in 24 manufacturing industries from 1948 to 1959. It defines and then tests three major hypotheses which are prevalent in the literature and seek to explain the cross-section differences and similarities among industries in amount of wage changes: (1) the "market hypothesis" (that "market forces" such as profit rates, rates of productivity increase, and employment changes are the main determinants), (2) the "union power hypothesis" (that degree of unionization has a noticeable and independent effect), and (3) the "national pattern bargaining hypothesis" (that the influence of a few "key bargainers" and political and economic pressures to imitate their actions are dominant).

Using average cents-per-hour wage rate increases, the tests find the national bargaining pattern to be influential over a certain limited area, but not an important independent force, since its influence depends on continued rough coincidence of market forces. Unionism was found to have a significant effect, even net of the various market forces variables. Market variables were found to be only weakly related to differential rate changes, perhaps because of the difficultly of measuring them meaningfully. Hourly earnings changes (which often differed considerably in amount from rate changes) were not well explained by pattern bargaining, were also related to unionism, and were a bit more closely correlated with economic factors (especially profits) than rate changes.

This essay was submitted to Yale University in 1962 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics at Yale and on the staff of the Economic Growth Center.

Dissertation Committee: Professors Mark W. Leiserson (Chairman), E. Wight Bakke, and Charles E. Lindblom.

*

The dissertation from which this essay was drawn was written under the supervision of Mark W. Leiserson and E. Wight Bakke, whom I wish to thank for their generous contribution of aid and encouragement. The subsequent comments of Charles E. Lindblom on both dissertation and essay were also most helpful and are greatly appreciated. The responsibility for remaining errors is, of course, my own.

D.R.S.

WAGE CHANGES IN 24 MANUFACTURING INDUSTRIES, 1948-59: A COMPARATIVE ANALYSIS

Donald R. Snodgrass

1. Hypotheses of Wage Determination

The period since World War II ended has been an interesting one for students of wage determination. It has seen rapid increases in money wages in most parts of the economy (but much faster increases in some industries than others) and a considerable volume of discussion of these phenomena in both the professional and the popular spheres. Among professionals there has been widespread disagreement over the choice of a theoretical apparatus to explain observed behavior. Most simply put, the debate is between those who feel that conventional economic theory can do a good job of interpreting the events of the last few years and those who feel that it cannot, but prefer to focus their attention on particular institutional arrangements (most notably the existence of the labor union) rather than talk about labor markets. This debate is of course of very long standing, going back to the earliest days of labor economics. In the specific context of postwar wage changes, it is reflected, for instance, in attempts to explain wage changes primarily by reference to the general level of unemployment,[1] union power in particular industries,[2] or the pervasive influence of a key bargain struck in one prominent bargaining situation.[3] Other writers have stressed the impact of still other institutional arrangements: e.g. long-term contracts[4] and industry-wide bargaining.[5]

1. William G. Bowen, Wage Behavior in the Postwar Period (Princeton, N.J.: Princeton University, Industrial Relations Section, 1960).

2. A. M. Ross, "The Influence of Unionism on Earnings," Quarterly Journal of Economics (February 1958), pp. 263-86; Arthur M. Ross and William Goldner, "Forces Affecting the Inter-Industry Wage Structure," Quarterly Journal of Economics (May 1950), pp. 254-81; Robert Ozanne, "The Impact of Unions on Wage Levels and Distribution," Quarterly Journal of Economics (May 1959), pp. 166-97.

3. Benson Soffer, "The Effects of Recent Long-Term Agreements on General Wage-Level Movements," Quarterly Journal of Economics (February 1959), pp. 36-60; ——, "On Union Rivalries and the Minimum Differentiation of Wage Patterns," Review of Economics and Statistics (February 1959), pp. 53-60; John E. Maher, "The Wage Pattern in the United States, 1946-57," Industrial and Labor Relations Review (October 1961), pp. 3-20; ——, "An Index of Wage Rates for Selected Industries, 1946-1957," Review of Economics and Statistics (August 1961), pp. 277-82.

4. Benson Soffer, "The Effects of Cost-of-Living Wage Escalators on General Wage-Level Movements: January, 1950 to June, 1953," unpublished dissertation, Princeton University, 1956.

5. Tipton R. Snavely, "The Impact of Multi-Unit Bargaining on the Economy," Southern Economic Journal (April 1953), pp. 445-57.

DONALD R. SNODGRASS

The aims of this paper are to be the main contending hypotheses of wage determination which are currently prevalent and to test them with reference to wage changes in 24 American manufacturing industries over the 1948-59 period The broad hypotheses which can claim applicability to this problem appear to be three in number. They will be referred to here as the market hypothesis, the union power hypothesis, and the national pattern bargaining hypothesis. This first section of the paper will go to some pains to define the hypotheses precisely and try to mold them into a form that will be empirically testable.

The Market Hypothesis

The "competive" hypothesis of wage determination is familiar to every economist and has always held a strong appeal within the profession for this very reason. Proponents of this thesis take the view that there is a labor market, which is much like any other essentially competitive market encountered in economic theory, and that the forces of supply and demand operate in this market to set a wage, which is a price like any other price. Probably in no other field where it has been applied has microeconomic theory been subjected to so many qualifications. Much of the empirical work of labor economics has dealt with particular qualifications which, the writers claimed, should be made. However, here as elsewhere, ". . . it takes a theory to kill a theory; facts can only dent the theorist's hide"[1] and many would still contend that the basic assertion of the competitive hypothesis is still the truest and broadest generalization that can be made about wages. This assertion is that "the market" (i.e. the joint actions of all workers and employers, with none separately having a perceptible influence) determines wages and that employers decide, on the basis of the marginal productivity criterion, how much labor to hire at these rates. All of this is viewed as a continuous process, with workers and employers reacting quickly to changes in market pressures.

The competitive hypothesis is thus completely straightforward theoretically. Making it testable, though, is very hard, in fact, in a literal sense, impossible. In past analysis, the most popular approaches to the problem have been two in number: (1) the Phillips[2] approach (employed in this country by Bowen[3]), which tries to relate variations in the rate of increase through time to over-all labor market supply and demand conditions (represented by the general level of unemployment) and (2) the cross-section approach of Dunlop and his successors,[4] which tries to relate interindustry differences in wage increase

1. Paul A. Samuelson, "Economic Theory and Wages," in The Impact of the Union, David McCord Wright, ed. (New York: Kelley & Millman, 1956), p. 323.

2. A. W. Phillips, "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957," Economica (November 1958), pp. 283-300.

3. Op. cit.

4. John Dunlop, "Productivity and the Wage Structure," in Income, Employment, and Public Policy: Essays in Honor of Alvin H. Hansen (New York: Norton, 1948), pp. 341-62.

to some factor which in some general way is related to market demand—rate of productivity increase, profit rate, employment change, or output change. Only the latter approach is involved here; our interest is exclusively with interindustry comparison. No completely satisfying explanations of why the competitive hypothesis would lead one to expect any of these quantities to be perfectly correlated with wage increases can be given. There is, however, some reason for expecting the correlations to be positive and fairly strong. High profits, rapid productivity increase, and output and employment gains would all be expected to be associated with rising labor demand, even though they are not identical with it. Separately or together these "market factors" can reasonably be expected to reflect at least some of the market pressures on wages. If nothing else, they can serve as standards against which the impact of other forces can be judged. Further discussion of the "market forces" variables can be found below (pp. 188–91), along with an historical review of their development in the literature and tests of their potency in explaining wage changes in the period of study. As is pointed out there, the looseness of the logical relation between the empirical measures and the competitive hypothesis, as well as the addition of the degree of product market concentration as an explanatory variable, make it not the competitive hypothesis itself but a more general, imperfectly competitive "market hypothesis" which is being tested in this paper.

The Union Power Hypothesis

The alternative theoretical strain to market theory is the notion that wages are the outcome of a bargaining process, a power struggle in which the determination and staying power of the employer and the employee (or his representative, the union) are tested. The notion of "bargaining power" has, however, remained hard to define in an operational manner (among other things, it is certainly related to market forces) and adequate analytical models of the bargaining process have been hard to formulate. The most important variant of the bargaining theory is the simple notion that "union power" can force wages above their market level. The era of precise statistical testing has revealed that this contention, still strongly made by most nonprofessional and some professional students of wages, is very hard to prove. Recent representatives of attempts to establish the hypothesis via serious empirical study have been Ross and Ozanne, the former using a cross-section analytical technique and the latter a time study approach. [1]

Two difficulties are commonly encountered in trying to prove the union power hypothesis. One is that union power, even if it does obtain higher wage increases than the market provides, may also induce higher wages for other units and its differential result is thus not measurable. The other is that even when unionized firms and industries have in fact gained more rapid rate increases, it has been hard to succeed in ruling out alternative explanations and parallel causation and establish an "independent" and "significant" impact

1. Op. cit.

for unionism. The attempts of Ross and Ozanne to overcome these difficulties are described below, pp. 184–86.

The approach to the problem of measuring "bargaining power" that is taken here involves the use of a figure representing the percentage of an industry covered by union contract. This, of course, fails to account for the "hardness of the rock through which the union must tunnel" [1] and does not really measure its proficiency at tunneling, either. Still, as empirical measures go in this field, it is a good one. Numerical percentage "union strength" is certainly an important consideration in over-all "bargaining strength."

The National Pattern Bargaining Hypothesis

The third major hypothesis to be considered here is that the impact of a very few "key bargains"—those between U. S. Steel (USS) and the United Steelworkers (USW) and the United Auto Workers (UAW) and General Motors (GM) are most often cited—on wages elsewhere is the dominant force in recent wage change experience. Every newspaper editorial writer appears to hold this view and so, apparently, does the President of the United States. These nonprofessional observers have been joined by at least a few professional students of wages, [2] who have tried to specify just what is meant by the hypothesis and see how well it stands up against the evidence. Since the hypothesis is the newest and perhaps least familiar of the three considered here, it deserves more detailed exposition.

The national pattern bargaining hypothesis represents a transference to the national scene of the notion of pattern bargaining, which has been recognized for some time to exist in a number of specific industries and localities. A pattern can be said to consist of a group of wage-setting units (firms, plants, industries, whatever the relevant unit is), a leader and a number of followers. The essence of pattern bargaining is that the actions of the followers are basically reactions to those of the leader. In the wage sphere, this can be further taken to mean that very nearly equal wage rate increases to those of the leader would be granted by the followers. [3]

How strong a tendency toward equality is it reasonable to expect? Certainly the hypothesis does not require that every single wage rate change of the leader be immediately and identically enacted by all the followers. Soffer has suggested that for pattern bargaining to hold the increases of the pattern participants cannot differ from one another by more than a random amount. [4] The use of the term "random" in this connection is somewhat irregular. It means that over time the cumulated wage increases of the follower units must show no

1. See Ozanne, pp. 182–83, for the origin of this analogy.

2. See note 3, p. 171.

3. See below, pp. 177–78.

4. Soffer, "The Effects of Cost-of-Living Wage Escalators . . . ," op. cit. p. 22.

consistent tendency to diverge in amount of cumulated increase from the cumulated value of the leader's increases. This stipulation makes the national pattern bargaining hypothesis testable: the participants in a national bargaining pattern must show no consistent divergence among themselves in cumulated cents per hour wage increase.

It should be stated, though, that even a finding of complete equality of cumulated increase among a group of units is not in itself irrefutable proof of pattern bargaining. The equality could be purely coincidental or it could be attributable to the incidence of similar forces other than the pattern. A priori knowledge of whether a pattern is likely to exist among the units in question is needed. Whether this likelihood is present or not depends on whether certain ties exist among the units, which serve as means of transmission for the key bargain. These ties are interrelations among units in regard to either markets or institutions. The market case involves a leader and followers who hire in a common labor market. Here, a given wage increase granted by the leader might well induce the followers to grant similar increases so as to retain their present labor force or their power to attract new labor. The institutional ties are those which make units part of an "orbit of coercive comparison,"[1] putting political pressure on the union officials and management of the follower units to imitate the key bargain. Specific ties of this sort include (1) common corporate ownership, (2) an oligopolistic leader-follower pattern in price setting and other nonwage matters, (3) a common union, and (4) a history of cooperation or (alternatively) sharp rivalry between the unions involved. Among units possessing either market or institutional ties of these kinds, equality of rate increases can be regarded as a likely indication of pattern bargaining.

So far nothing has been said about how the key bargain itself is reached: indeed, this is usually neglected in studies of pattern bargaining. For the key bargainers there are two basic possibilities: supply and demand or union power. Thus, the national pattern bargaining hypothesis is really a special case of the other two hypotheses. It is significant, though, that it is the supply and demand or bargaining power of the key bargainers that matters here, not those of the great number of units. Should the national pattern bargaining hypothesis prove to hold any substantial independent significance, both the question of its transmission mechanism and that of the origin of the key bargain will have to be studied. The three major hypotheses of wage determination and their variations are outlined below in Table 1.

In regard to each of the hypotheses, it will be asked how great the influence of the determinants it specifies seems to be. With national pattern bargaining hypothesis, though, the special question of the range of its application also seems pertinent. Although its proponents do not always make the point clear, it seems impossible that the hypothesis as defined above (pp. 174-75 could be meant to apply to wage changes in the whole economy. The great

1. This term was first used by Ross. See Arthur M. Ross, Trade Union Wage Policy (Berkeley and Los Angeles: University of California Press, 1948), pp. 53ff.

bulk of evidence adduced in its behalf (for instance, by Maher)[1] covers only production workers and mainly the durable goods manufacturing industries. Little or nothing is said about its impact on white collar employees, soft goods industries, and nonmanufacturing. Some "ripple" effect of spreading but weakening influence is doubtless envisioned. The generality of the hypothesis in the form in which it has been stated above (which demands a long run tendency toward equal wage increases) is thus bound to be limited.

TABLE 1

The Hypotheses of Wage Determination

	Wage Hypotheses		
Type of firm	The Market	Union Power	Pattern Bargaining
"Ordinary firms"	1. Raise wages in accord with supply and demand.	2. Raise wages as a result of union bargaining power (vis-a-vis that of employer).	3. Raise wages in imitation of a leader. a. Leader influences supply and demand of ordinary firms. b. Leader influences bargaining power in ordinary firms.
Actual or potential "key bargainer"	1. Raise wages in accord with supply and demand.	2. Raise wages as a result of union bargaining power (vis-a-vis that of employer).	3. Excluded by definition.

A suggested qualification of the national pattern bargaining hypothesis has been put forth by Soffer.[2] He suggests that there is an "auto-steel orbit" of industries which possess the requisite ties of a bargaining pattern with the key bargainers in the auto and steel industries. Thus General Motors has ownership ties in the electrical equipment industry, the United Auto Workers are also active in the farm machinery and aircraft fields, the International Association of Machinists rivals the UAW in several areas, etc. Many such

1. "The Wage Pattern in the United States . . . ," op. cit.

2. Soffer, "The Effects of Cost-of-Living Wage Escalators . . . ," op. cit., pp. 99 ff.

potential transmission links could be mentioned. This qualified version of the national pattern bargaining hypothesis deserves to be given some consideration. As a more realistic version of the hypothesis, it will be of some interest, even though it forfeits generality by failing to specify how wages are determined in the rest of the economy.

These are the rival hypotheses as they will be used in this paper. As a first approximation, they will be viewed as competing with each other. There is, though, no logical necessity that they be mutually exclusive. The key bargain could be closely in line with supply and demand conditions (but only if these are quite uniform as among units); the view that unions raise wages does not rule out a possible market impact; and so on. After testing them as alternatives, ways will be sought to reconcile them as seems necessary.

The Wage Variable

Before tests are performed, though, a question must be considered which has been avoided up to now. To what particular wage measure do the hypotheses refer? The market hypothesis talks about the "price of labor." The union power hypothesis refers to the power impact on wage rates, but could also be broadened to the idea that unions gain a wide variety of monetary benefits for their members. What the national pattern bargaining hypothesis refers to is not clear; most literally, it is probably to whole "packages" of contract benefits. If the hypotheses are to be compared, though, a common wage measure must be agreed upon. By all odds the most generally satisfactory one is average money wage rates.

Wage rates are very near to the "price of labor" concept. As compared with their main rival in empirical work, average hourly earnings, they are a "purer" price concept, lacking the large quantity admixture of the latter measure. They are also likely to be a major, though not the exclusive concern of both union power and pattern bargaining. The reason for this is a political one. Collective agreement packages, whether they are imitative of a key bargain or not, contain a number of diverse kinds of benefits. Some of these are incapable of translation into money terms; some can be so translated, but only on rather arbitrary assumptions about how much the provision will come into actual play; and some are inherently nonmonetary, granting privileges instead of pecuniary benefits. In this maze of clauses, one consideration, the amount of wage rate increase, stands out as the best single quantitative indicator of the magnitude of the settlement. It receives the most prominence in the newspapers and probably features most prominently in the minds of the union members, too. It is thus extremely important to the union. It forms the most obvious measure of its tangible attainments, quantifying its ability to get "more" or "as much as the other fellow," which ever rationale applies. Thus, to "keep up" with a leader it would normally be necessary to obtain a wage rate increase of a very similar amount to that of the leader.

Not only for the union member but for the interpreter, too, the amount of wage rate increase serves as an indicator of the size of the total package.

One's expectation would normally be that a positive relation exists between high wage rate increases and liberal fringe benefit improvements; the same firms and industries would be expected to give both. True, there is a counter tendency, noted by Soffer.[1] He argues that it is a common practice among pattern bargainers to engage in a "minimum differentiation" of the form of the leader's bargain, while retaining its over-all amount, in order to give a spurious impression of independence. A really strong tendency for wages and fringe benefits to serve as substitutes for each other within uniform packages would make it unreasonable to expect any tendency toward uniformity to emerge in statistics of wage rate change. However, Soffer offers no evidence on the prevalence of minimum differentiation and it seems likely that the complementarity effect would swamp the substitutibility effect statistically in any given body of data, so this possible complication will be ignored. We may thus be reassured that the national pattern bargaining hypothesis implies a tendency toward equality in wage rate increases among participating units.

As the best compromise to fit all three hypotheses, wage rates will be used as the main dependent variable in this study. Whole benefits packages would also be interesting to test, especially in regard to the union power and national pattern bargaining hypotheses, but they are inherently hard to quantify and no standard data on them exist. The advisability of testing the hypotheses with average hourly earnings, though, should not be ruled out. For one thing, past studies of the market and union power hypotheses have employed earnings and parallel tests are thus needed for comparision. Also, earnings are important in their own right and versions of the hypotheses which have been restated to apply to earnings rather than rates are only slightly less interesting than the wage rate versions.

Collective bargains and similar decision-embodying documents set wage rates. These documents are drawn up periodically and are intended to cover a fairly fixed future period of time (this is especially true of collective bargaining, but may also apply in the nonunion case, if only because of the time required for calculation and decision making). They actually set up a structure of rates, applying to all the possible circumstances under which work may be done in the future; there are straight time occupational rates, overtime rates, special shift rates, and so on. There are also nonrate provisions which help to govern the applicability of the various rates and thus affect future earnings. Earnings over the period in question, though, are not entirely determined at this time, but also depend on the set of circumstances that actually materializes in the future, on the reactions of the parties to the circumstances, and on any additional earnings-influencing activities on the parties. Many different kinds of actions and reactions—on employment, work scheduling, administration of incentive payment systems, job evaluation, etc.—affect earnings. Some are created by changing market circumstances, some by union pressure, and some by management initiative. Logically, then, it is evident that the conclusions reached on the market, union, or pattern

1. "On Union Rivalries . . . ," op. cit.

bargaining impact on hourly earnings may well be different from those reached on rates.

Moreover, earnings are theoretically important, more so than has been implied so far. They are average "price of labor" which corresponds to the wage rate's position as the marginal price of labor in particular circumstances. Thus, from the point of view of labor cost, the wage rate is the relevant variable in a very short run marginal employment decision (e.g. whether to work a particular employee an hour of overtime) but average hourly earnings are the pertinent measure in longer run decisions, such as whether to accept a given production contract. Who can say that only the former type of problem is of theoretical interest? The issue turns on the time horizon over which managers are thought to view wage-employment decisions.

The macroeconomic importance of the level of average earnings in a large sector of the economy is so obvious as to require little defense. The earnings concept, when applied over a broad area, corresponds with the notion of a wage level, as it is used in aggregate economic theory. It is a measure of wage cost per hour of labor input and, with productivity, of wage cost per unit of output. On the demand side, it is, with total man-hours, an important element of income and aggregate demand.

Each of the three hypotheses can be restated in such a way as to make it apply to earnings rather than rate changes. Such a restatement of the national pattern bargaining hypothesis would involve dropping an assumption made earlier (see pp. 177–78, above): that imitation of the key bargain necessarily implies a tendency toward the equalization of cumulated wage rate changes. Now the area of imitation can be broadened to include all earnings-increasing provisions of collective bargains. The hypothesis now is that pattern bargaining packages have an approximately equal impact on earnings levels. It is not reasonable here to expect more than an approximately equality of earnings increases in two units would require not only identical agreements (in the absence of a fluke coincidence) but also the identity of the environmental and administrative experience of the units over the life of their agreements. Still, it will be interesting and meaningful to see whether the auto-steel orbit grouping of industries is more or less easily distinguished from the other industries in the study as a result of the shift from rate to earnings changes.

A similar adjustment can be made in the union power hypothesis. It will now be taken to imply that unions have a significant impact not just on rates but on all earnings-increasing provisions of labor contracts. It may also be taken to imply that unions may raise earnings levels through their activities between bargaining sessions, through grievance procedures and other channels for influencing wage administration.

These two hypotheses assert an essential parallelism between the forces which create wage rate increases and those which influence earnings levels. One interpretation of the market hypothesis which can be taken from the past literature on wage determination leads one to believe that the market can be a more potent force in the determination earnings levels than in the

setting of rates. For the postwar United States, deviations between rate and earnings movements have received virtually no attention. They have, however, been closely scrutinized in recent years in certain European countries (notably the United Kingdom and the Scandinavian countries) and they were also studied in connection with the World War II American economy.[1] A common line of reasoning can be traced through all of these studies. It is that wage rates are set by a rigidly formal institutional process (in the cases in question, national wage control or a voluntary form of control, "wage restraint") which does not allow them to find their market equilibrium level. (In the specific cases of the literature in question, rates were, of course, kept below their equilibrium level.) Market forces (in these cases excess labor demand) then pulled earnings more into line with supply and demand conditions. The upward deviation of earnings from rates in Scandinavia and the U.K. has been labeled "wage drift" and traced to the existence of a number of mechanisms (ingrade progression, hiring at higher rates than usual, loose administration of piece rate systems, and others) which allow the payment of average earnings which are higher than the officially scheduled rates. A generalization of this experience would involve two assertions: that the price of labor is mainly market-determined and that, because of institutional rigidities, earnings are a better measure of the true market price of labor than are rates. Assuming that this line of reasoning is reversible (i.e. that it applies in cases of deficient as well as excess labor demand at scheduled rates), one would be led to expect better correlations with market factors when earnings are used than with rates.

However, it must be recognized that a great many "mechanisms" of earnings drift are possible (see pp. 178–79 above) and that in the absence of detailed empirical study it is impossible to say which ones in fact accounted for the earnings-rate deviations observed in the period and industries studied here. They may or may not have been market-related, as the hypothesis just outlined implies they were. A possible outcome of the tests to be conducted here, then, is that only the "noise" in the correlations will be increased. If

1. B. C. Roberts, National Wages Policy in War and Peace (London: Allen and Unwin, 1958), discusses several experiences with a national wage policy. See also, for the World War II American case, John T. Dunlop, "An Appraisal of Wage Stabilization Policies," in Problems and Policies of Dispute Settlement during World War II (Washington: Department of Labor, Bulletin 1009, 1947), pp. 536–59. For West European experiences, see a host of works including Bent Hansen and Gosta Rehn, "On Wage Drift," in Twenty-Five Economic Essays in Honour of Erik Lindahl (Stockholm: Ekonomisk Tidskrift, 1956), pp. 87–138; Mark W. Leiserson, Wages and Economic Control in Norway (Cambridge, Mass.: Harvard University Press, 1959); T. L. Johnston, "Wages Policy in Sweden," Economica (August 1958), pp. 213–29; H. A. Turner, "Wages: Industry Rates, Workplace Rates and the Wage Drift," Manchester School (May 1956), pp. 95–123; and—, "Wages, Productivity and the Level of Employment: More on the Wage Drift," Manchester School (January 1960), pp. 89–123.

the mechanisms of earnings drift are not clearly related to any of the three hypotheses, all the coefficients will fall in value and only the unexplained variation will be higher.

The Research Program

We now have assembled three major hypotheses of wage determination, a number of variants, and two sets of wage change data on which to test them. The tests involve changes in the wage rates and hourly earnings of the production workers of 24 manufacturing industries over the years 1948–59. They specifically ask: how do the hypotheses fare as explanations of the relative amounts of wage increase recorded by the 24 industries? Their limited direct application is evident. Extrapolation of findings to other times and places will be risky indeed.

2. Wage Rate Changes and Related Variables

The Data

Data on absolute wage rate changes for 24 manufacturing industries are presented below in Table 2. These are cents per hour changes over the years 1948–59 inclusive. Since the collective bargaining of the late 1940's was conducted in a number of fairly discrete "rounds," it is essential to begin at the start of a round and not break into the middle of one. The beginning point of this study, January 1948, coincides with the start of the third postwar round; the years of the first two rounds, 1946 and 1947, are excluded from this study because they are not covered in the primary source of wage rate data. By the end of the twelve-year period, the division of wage actions into neat rounds had almost disappeared, so the choice of a cut-off point is less crucial. Fairly comparable periods for the various industries should be the result (though steel, as will be seen, is one industry which is understated as a result of the choice of December 31, 1959, as the cut-off date).

The data of Table 2 were compiled with the use of a number of sources which report on current wage rate changes.[1] These figures represent currently weighted (by employment) averages for a sample of firms in the various industries. Four specific points on their compilation deserve mention:

(1) Industries with the most wage rate data were chosen for the study, but no rigid standard on minimum coverage was adopted. Typically half the units in the industry were included in figuring the average wage rate change for any given year, but degree of coverage is not in any case wholly adequate as a test of reliability, since degree of variation in rate changes among firms is also important. A 20 per cent sample in some industries may well be more accurate than a 60 per cent sample in others.

(2) The primary wage announcement data always refer to particular instances of rate change and not to the total of such changes over the year. It may be that some firms which have made more than one change during the year

1. See "Sources," Table 2, p. 182.

TABLE 2

Cumulated Wage Rate Changes in 24 Manufacturing Industries, 1948-59 Inclusive [1]

(cents per hour)

Industry	Wage Rate Changes	Industry	Wage Rate Changes
Flat glass	88.5	Meat packing	117
Glassware	91.5	Canning	78
Cement	115.5	Grain milling	101
Basic steel	117.5	Baking	105.5
Ferrous foundries	100	Sugar	107.5
Nonferrous metals	113	Textiles	52
Tin cans	112	Apparel	55.5
Farm machinery	110.5	Paper	86
Electrical equipment	99.5	Chemicals	102
Automobiles	111	Petroleum refining	111
Aircraft	113	Rubber	106.5
Shipbuilding	119.5	Shoes	48.5

1. See text, pp. 181-83, for details on the compilation of these figures.

Sources: The main source was Current Wage Developments, a monthly publication of the Department of Labor. Also used were the Bureau of Labor Statistics (BLS) Wage Chronology Series (installments of which appear from time to time in the Monthly Labor Review) and the independent listings of the National Industrial Conference Board, a monthly feature in the Management Record.

were not fully covered and that some which made no change at all were not counted as zeros, but were simply left out of the sample instead.

(3) Besides general rate changes, some of the more important special changes (for all women, all skilled workers, a given region) were included, while those applying to still narrower groups were not.

(4) Rough estimates of a few changes had to be made where rate changes were reported in percentage terms and the precise average rate base was not known and where the exact size of a subgroup of workers was not known.

For all these reasons, no inordinate degree of precision should be attributed to the wage rate measure. Neither the magnitude nor the direction of error is entirely clear. More likely than not, the direction of any bias is upward. In many industries the figures tend to be a rather complete account of the activities of the larger firms in the industry. Certainly zero increases have escaped inclusion more frequently than positive ones. An interpretation of the estimates as the maximum possible values which actual wage rate

increases could have assumed would not go too far wrong. As for the size of error, all that can be suggested is that it is almost certainly small enough to make this somewhat rough but direct estimate of wage rate change a considerably closer approximation of its true value than average hourly earnings, the frequently employed substitute for wage rate data.

The National Pattern Bargaining Hypothesis

Maher has studied in detail the actions of the most prominent rate setters in the economy over the 1946–57 period.[1] His work cites the automobile and steel bargainers as the source of the key bargain which influences the actions of the participants in a national bargaining pattern. The United Auto Workers–General Motors and U.S. Steel–United Steelworkers are seen as the main initiators of the key bargain, with detailed study of bargaining history being counted on to reveal which has served as the key bargainer at any particular time. Following Maher's methods, it is possible to specify a cumulated key bargain wage rate increase of $1.135 for the 1948–59 period under study here.[2] This is the standard against which potential participants in the national bargaining pattern must be measured. How dominant and how widespread is its influence? The simplest kind of comparison of this figure with the increases for the industries listed in Table 2 reveals that some of these industries gave cumulated increases of about the amount of the cumulated key bargain, but many others did not. In its most literal sense, that all manufacturing industries gave nearly identical cumulated increases, the national pattern bargaining hypothesis clearly does not stand up to the evidence.

It may be, though, that the key bargain has a substantial impact in a segment of the economy, perhaps a fairly wide one. The auto-steel orbit variant of the hypothesis divides the economy into two groups of industries, one essentially a part of the national pattern and the other made up of units which cannot or need not join in the pattern. The auto-steel orbit industries are said to be those which have market or institutional ties to the auto and steel key bargainers (see pp. 176–77, above). Of the industries under study here, eleven can be considered to possess such ties: flat glass, basic steel, ferrous foundries, nonferrous metals, tin cans, farm machinery, electrical equipment, automobiles, aircraft, shipbuilding, and rubber. Using this division, it is found that the mean value of cumulated rate increases within the orbit is 109.2 cents, while the mean value for the other industries is only 90.1 cents. An "F" test for variance reveals that the former figure is significantly nearer the

1. "The Wage Pattern in the United States . . . ," op. cit.

2. Maher's figure for the annual key bargain increase is the one used for the years through 1957. For the last two years of this study, which were not covered by Maher, Maher's method—which consists of identifying the key bargain ex post on the basis of which of the leading candidates for that position seems to attract the most followers at any point in time—was used to identify the key bargain.

cumulated key bargain total of $1.135 than the latter, with a probability of .95.[1] The applicability of a national pattern bargaining hypothesis to a wide range of the economy seems vindicated.

But is this the most revealing way of viewing wage determination? Doubtless the specification of an orbit of politically and economically interrelated industries reveals something about the data, but the importance of such a set of ties relative to other influences on wages cannot be regarded as having been established here. Associative statistical analysis never proves causation and other hypotheses may fit the data as well or better than the auto-steel orbit variant of the national pattern bargaining hypothesis does. For instance, an alternative way of dividing up the industries of Table 2 is to distinguish between industries which bargain for wages on an essentially centralized basis and those which do not. A plausible case can be made for this distinction on the grounds that centralized bargains are more likely to follow the national key bargain because their scope is essentially national and they are less involved with local variations in market and political considerations. The same kind of "F" test applied to this division of industries shows it also to be statistically significant.[2]

Thus the acceptability of a qualified national pattern bargaining hypothesis in either of two variants seems superficially to be indicated, but this only complicates matters. For one thing, the two groupings strongly coincide in make-up. More important, the pattern bargaining sector tends to be the high-wage, durable-good, highly concentrated, heavily unionized, profitable, big-city oriented sectors also. With such a tangle of possible causal forces, how can a final conclusion on the national pattern bargaining hypothesis ever be reached?

The best way out of this difficulty seems to be to proceed to the testing of alternative explanations of wage determination. How much of the observed behavior of rate changes can be explained totally without resort to the notion of a national bargaining pattern? Should the amount turn out to be great, no clear-cut basis for choosing among hypotheses will have emerged from the data. On the other hand, external logic may give some assistance in making a choice among apparently equally tenable hypotheses.

The Union Power Hypothesis

Two studies may be selected as typical of past attempts to show the importance of union bargaining power in wage determination. One was by

---

1. $F = 6.24$, as compared with $F_{.95} = 4.28$ and $F_{.99} = 7.88$.

2. If the criterion is used that centralized bargaining exists where wage rates are set for a majority of workers in the industry by three or fewer bargains, then the flat glass, basic steel, nonferrous metals, tin cans, farm machinery, electrical equipment, automobiles, shipbuilding, meat packing, and rubber make up the centralized bargaining group. Here, $F = 6.40$.

Ross.[1] Ross studied percentage increases in straight-time average hourly earnings, 1933–45, in 44 manufacturing and extractive industries and came to the conclusion that in low-, medium-, and high-wage industries alike there was a clear-cut positive relation between these increases and the percentage of the industry labor force under union contract in 1945. Several alternative explanations of these findings (changes in skill levels, rural versus urban location, proportion of women employed, degree of monopoly, and direction and rate of employment change) were considered by Ross, but he concluded that ". . . upon examination none of these factors will serve to explain away the evident association between unionism and earnings."[2]

Ross was soon to qualify his certainty on this point. Just two years later, in an article coauthored with Goldner, he retreated to a position that new unionization, though not necessarily old unionism, had a positive influence on earnings.[3] The reason for Ross' retreat was the strong contrary arguments of several writers whose main offensive weapon was the contention that coinciding factors, not considered by Ross, could explain his results.[4] Among such factors cited by Dunlop were proportion of labor costs in total costs, degree of product market monopoly, output and productivity increases, and changes in skill and other occupation mixes.[5] In their answer to the critics, Ross and Goldner admitted at least employment change and product market concentration as additional factors influencing earnings. They restated their conclusion on unionism to read: "Among industries which were substantially unorganized in 1933, subsequent increases in earnings were associated with changes in the degree of organization. However, those who were already substantially organized in 1933 have lagged behind all other groups."[6]

The proposition that the influence of unionism on wages is, at the very least, hard to establish because of the simultaneous existence of contending explanations seems to have remained the received doctrine in labor economics since 1950. No cross-section study made since that time has given a major place to unionism and other factors have been given more stress (see, for instance, the studies reviewed below, pp. 187–90). Recently, though, a paper by Ozanne reopened the possibility of a systematic gain from unionism in the

---

1. "The Influence of Unionism on Wages," op. cit. (reprinted in Trade Union Wage Policy, op. cit.).

2. Ibid., p. 128.

3. Ross and Goldner, op. cit.

4. C. L. Christenson, review of Trade Union Wage Policy, Journal of Political Economy (February 1949), p. 86; Richard A. Lester, "Comment," Quarterly Journal of Economics (November 1948), pp. 783–91; and Dunlop, op. cit.

5. "Productivity and the Wage Structure," op. cit., pp. 351–54.

6. Op. cit., p. 280.

postwar years.[1] Ozanne's analysis was based not on the traditional cross-section method but on a time study approach. Criticizing the cross-section method for its inability to distinguish among variables or measure the induced effects of unionism in one unit or sector on wages elsewhere he shifted to comparison of two apparently similar periods in American economic history: 1923-29 and 1947-56. Finding a considerably greater rise in the money wage level over the latter period, he attributed it to the greater prevalence of unionism. He found no reason among market considerations for the difference between periods. Still, it is difficult to believe in the possibility of adequate statistical control for "other things" in comparing two historical periods, and Ozanne's analysis, though interesting, cannot be considered to have ended the debate over wages and union bargaining strength.

Ozanne ended his article with the following prediction: ". . . I would suggest that the accumulating bulk of evidence gathered since 1947 will support the notion of businessmen, politicians, and trade unionists—that collective bargaining is more than a ritual."[2] The time has come to look at this accumulating evidence.

Measuring union strength by percentage of industry production worker employment under union contract and continuing to measure wage rate changes in cents-per-hour terms, a close correlation is indeed uncovered. A simple rank correlation of +.737 is found, which is significant with .99 confidence.[3] Because of the well known and often cited possibility of parallel causation, nothing conclusive is proven by this simple correlation. All that can be said at this juncture is that the union power hypothesis is another possible but not definitely established explanation of wage rate changes, 1948-59.

One interesting point about this finding should be made, though: it conflicts with the conclusions of the two earlier cross-section analyses of similar post-World War II time periods (those of Bowen and Levinson).[4] Neither of these found unionism to be a significant explanatory factor. Levinson found no rank correlation over the 1947-58 period between percentage change in straight-time earnings and 1946 unionization data[5] for 20 S.I.C.

---

1. Op. cit.

2. Op. cit., p. 196.

3. If the less defensible normal correlation coefficient is used, a slightly higher value, +.775, results. The 24 industries are by no means normally distributed with respect to unionism, though, so less reliance can be put in this measure. See Appendix A, pp. 212-15, below, for a description of the data and a fuller discussion of the reliability of these statistical findings.

4. Harold M. Levinson, Postwar Movements of Prices and Wages in Manufacturing Industries (Washington: Joint Economic Committee, Study Paper 21, 1960); Bowen, op. cit.

5. See Appendix A, note 5, pp. 213-14.

two-digit manufacturing industries. Bowen used percentage changes in gross earnings and the 1958 unionization data employed here on all two-digit manufacturing industries over the time span of January 1947 to June 1959; he found unionism to be positively related to his dependent variable in only three of the six cyclically-chosen subperiods he identified within his over-all period (as a partial correlation coefficient in an equation which also included employment change, profits, and concentration as contributing factors to multiple causation).

How are the differences between the findings here and those of these two earlier studies to be reconciled? Both Bowen and Levinson use some form of average hourly earnings instead of wage rate changes. But this difference does not help to explain the discrepancy: substituting changes in gross earnings for rate changes in the present paper's group of industries and time period actually raises the rank correlation with unionism slightly, from +.737 to +.750. Nor is the particular group of industries selected for study here especially favorable to the case for a significant union impact. Substituting gross earnings changes for all two-digit industries as the dependent variable for the 1948–59 period again raises the coefficient of rank correlation, this time to +.802. The group of industries used here certainly does not exaggerate unionism's importance.

Closer scrutiny of the two studies shows the origin of the divergent conclusions to be relatively minor procedural differences. Levinson never formally ran correlations with unionism at all, but his data can, for his entire twelve-year period, be interpreted as giving a rank correlation of only +.129 with the 1946 unionization data he uses. If, however, a change is made to the 1958 data, the coefficient jumps to +.589[1] (this is on percentage changes; correcting the data of this paper to allow for differing original wage levels reconciles the remaining discrepancy). Bowen talked about subperiods of roughly two years' duration. He found no consistent relation with unionism, but then he was looking at a short-run case, a different one from the object of this and most other studies. In his case it is doubtful whether anyone would predict a consistently high correlation. If, however, Bowen's short periods are added up and viewed as one long twelve and a half year period a rank correlation of +.595 is the result. Thus, using the more recent data on unionism and taking wage changes as a trend over a postwar period of ten years or more, the findings of these interpreters are not inconsistent with the correlation presented here.

The Market Hypothesis

In a field of study where measurement problems abound, the competitive hypothesis has stoutly resisted any very direct empirical testing. Beginning with Dunlop's 1948 article, substitute measures of market forces have been suggested by various writers.[2] Dunlop's real-world stand-in for the

1. See Appendix A, note 5, pp. 213–14, for elucidation of this distinction.

2. "Productivity and the Wage Structure," op. cit.

competitive hypothesis was rate of change of average physical productivity. His suggestion that wage changes follow productivity changes was really put forth more as an ad hoc generalization than as a deduction from competitive theory, since nothing in the latter body of thought can be interpreted as implying that wages in a competitive labor market need behave this way. What might be called a pseudo-theoretical backing for the measure might be given by calling it a measure of "ability to pay," but since this is nothing but a nebulous term for market forces it is probably best simply to regard Dunlop's proposition as an empirical one and not try to carry its theoretical implications too far. What is being tested, both here and in the other studies that will be cited below, is thus not the competitive theory but a much looser assertion that market forces matter.

In his contention that a close empirical relation exists between productivity change and wage change, Dunlop recognized several complicating factors, while minimizing their net effect: the possibility of price cuts rather than wage increases as reactions to productivity gains, especially in highly competitive product markets; the need to preserve fairly equal pay for equal work among industries; and the bargaining power of unions. For the 1923–40 period which he studies he obtained a +.47 rank correlation and a +.60 normal correlation between average productivity changes and gross average hourly earnings changes and felt that his fundamental relationship was strongly borne out.

Several later writers worked at modifying Dunlop's train of thought, trying to take fuller account of the galaxy of possible market influences on wage determination. Garbarino built on Dunlop's work by pointing out that while productivity increases opened up the possibility of wage increases there was also a possibility that consumers, suppliers of nonlabor factors of production, and company owners might also benefit, depending (in Dunlop's own terminology [1]) on the structure of related markets.[2] Garbarino took the concentration ratio of the product market and the familiar unionization ratio as the relevant market structure measures and got, for the 1923–40 period, +.60 correlation between gross average hourly earnings and productivity, +.67 with concentration, and .856 with both taken together. The relation with unionism was weak.

Perlman tried to improve on Garbarino in restating Dunlop's thesis.[3] He suggested a substitute solution to the problem of alternative uses of productivity gains: a shift from physical to value productivity (from output to sales per man-hour). As a measure of "ability to pay" this had the advantage

1. See John T. Dunlop, Wage Determination under Trade Unions (2nd. edn., New York: Augustus M. Kelley, 1950), pp. 74ff.

2. Joseph W. Garbarino, "A Theory of Inter-Industry Wage Structure Variations," Quarterly Journal of Economics (May 1950), pp. 283–305.

3. Richard Perlman, "Value Productivity and the Interindustry Wage Structure," Industrial and Labor Relations Review (October 1956), pp. 26–39.

of allowing rises in product demand which expressed themselves in higher product prices but not in higher physical productivity to have an impact on wages. But it opened the door to reverse logic and doubtful conclusions, since if rate increases (however caused) themselves cause price increases and these raise the value of sales, value productivity will rise. A wage-induced rise in value productivity thus cannot be distinguished from a productivity-induced rise in wages. [1]

Other measures of market forces have commanded some popularity in empirical work (again, with relatively little formal theoretical support). Profit rates have been used by many writers and employment and output growth are other variables which have been used. All might be expected to have a loose connection with labor demand, but since their theoretical grounding is weak, a very straight forward attitude can be taken here. It will be sufficient for present purposes just to ask whether any of these potential independent variables, singly or in combination, provide an adequate explanation of the wage rate changes listed in Table 2.

Table 3 begins to provide an answer. As a simple rank or normal correlation coefficient, only the concentration ratio emerges as a significant factor in determining these rate changes. The relations with productivity change, profit rate, employment change, and output change are all weak. At this point an issue which has been avoided so far should be faced. Should wage rate changes be measured in percentage or absolute terms? There was a time when the former method would have been used without consideration of the alternative and, indeed, one's first inclination is to measure any change this way. Ross and Goldner, [2] though, challenged this procedure in connection with one aspect of their study, the question of whether interindustry wage differentials had widened or narrowed. Their grounds for choosing the absolute measure instead were that in the period they were studying diverse percentage changes obscured the existence of a strong and meaningful uniformity of absolute changes. The choice in any particular case must be made on grounds such as these; there is no a priori reason why one method or the other must be chosen in all cases. Here our interest centers around the national pattern bargaining hypothesis, which alleges uniformity in absolute changes, and alternatives to it, so the cents per hour absolute measure is appropriate. This is not to deny that the initial wage level has an impact on subsequent wage rate changes, as percentage measurement of changes implies. When using absolute changes one must guard against overlooking this impact. As the final column of Table 3 shows, when adjustment is made for initial wage level differences, all market forces indicators explain wage rate change even less well than before. Most notably, the seeming significance

1. With physical productivity, this is not much of an analytical problem; wage-induced rises in physical productivity are highly unlikely in the context of post-war America; see pp. 191–92, below.

2. Op. cit.

TABLE 3

Simple Correlations between Average Wage Rate Change in 24 Manufacturing Industries and Various Measures of "Market Forces," 1948–59 [1]

Independent variable	Rank correlation coefficient	Normal correlation coefficient	As a partial coefficient with the initial wage level
Productivity (change in output/production worker man-hour)	+.212	+.217	+.143
Profits (after-tax, as a percentage return on stockholders' equity)	.146	.331	.154
Employment change (as a percentage change, production-worker man-hours).	−.042	.141	.109
Output change	.066	.021	.004
Concentration (percentage of sales by four largest firms)	.488*	.444*	.228

1. For the basic data and a detailed description of their definitions and sources, see Appendix A, pp. 212–15.

*Coefficient is significant with .95 confidence.

of the relation between market concentration and wage change is eliminated if one adjusts for differences in original wage levels.

The obvious imperfection, both theoretically and as regards measurement, of the "market forces" might lead to some considerable skepticism as to whether the market hypothesis has received a fair test here. One thing that can be done to give it more of a chance is to combine variables into multiple correlations and see whether in concert they appear to explain more of the observed wage rate change than they do individually. The results of a series of such multiple correlation tests appear in Table 4. Again, the market forces indicators fail the test. In none of their possible combinations do they provide a statistically significant explanation of wage change.

TABLE 4

Coefficients of Multiple and Partial Correlation:
Wage Rate Change in 24 Manufacturing Industries
and Various Measures of "Market Forces," 1948-59 [1]

Equation	Partial correlation coefficients: Productivity	Profits	Concentration	Multiple correlation coefficient
a. [2]	+.195	+.318		.379
b.	.254		+.447*	.487
c.		.168	.348	.466
d.	.192	.107	.346	.496

1. Neither output change nor employment change is included in these multiple correlations, since both quickly became irrationally negative when combined with other independent variables.

2. Each equation combines the variables for which partial correlation coefficients are included, with the indicated results in partial and multiple correlations.

*Coefficient is significant with .95 confidence.

The Market and Union Power Combined

A further step which can be taken with this correlation analysis is to combine unionism with the economic variables. Such a test will examine the plausibility of a combined market and bargaining hypothesis. Such a dualistic mode of wage rate setting is, after all, a more realistic probability than either of the polar positions (that either supply and demand or bargaining power alone sets wages). In Table 5 this has been done in various combinations of variables. Here a certain degree of success is obtained. Either of the last two equations listed in Table 5 explains roughly two thirds of the total interindustry variation in wage rate change. Still very obvious in these equations, though, is the failure of any of the market forces indicators to become statistically significant at the .95 confidence level. Productivity change, profit rate, and concentration ratios all show consistently positive relations to wage change, as would be expected, but the strength of the relation never becomes very great.

If one were to allow the measure of value productivity, suggested by Perlman, [1] market forces would be made to look somewhat more influential

1. Op. cit.

TABLE 5

Coefficients of Multiple and Partial Correlation: Wage Rate Changes in 24 Manufacturing Industries and Various Combinations of the Independent Variables, 1948–59

Equation	Productivity	Profits	Concentration	Unionism	Multiple correlation coefficient
a.[1]	+.197			+.773**	.785**
b.		+.180		.752**	.787**
c.			+.309	.741**	.799**
d.	.174	.154		.750**	.791**
e.	.213		.319	.740**	.809**
f.		.144	.291	.740**	.804**
g.	.194	.112	.303	.734**	.812**
h.[2]	.192	.099	.250	.673**	.812**

1. Each equation combines the variables for which partial correlation coefficients are included, with the indicated results in partial and multiple correlations.

2. Equation h refers to a final formulation in which the initial wage level was included as a fifth independent variable. Its partial correlation coefficient was only +.005. It reduced the coefficients of the other variables by the amounts shown and left the coefficient of multiple correlation essentially unchanged.

**Coefficient is significant with .99 confidence.

relative to unionism. Perlman's measure (i.e. value of sales/man-hours of production-worker input) is what might be called gross value productivity. Its rank correlation coefficient with wage change is +.244 (in a normal equation form it achieves a correlation of +.512, which is statistically significant at the .99 confidence level). It records changes in value per man-hour remaining in the firm for division among ownership, labor, and other factor suppliers. The suppliers of materials to the firm would also be eliminated from the picture if value added were used rather than gross sales as the numerator of the value productivity fraction. This change greatly enhances the apparent explanatory value of productivity—the rank and normal simple correlation coefficients rise to +.674 and +.725 respectively, both clearly significant at the .99 level, and the over-all relationship improves, too (see Table 6).[1] But the value productivity concepts, as has already been

1. The net value productivity coefficients refer to the 1948–57 period. By way of contrast, the simple rank correlation of gross value productivity for 1948–57 was only +.204.

TABLE 6

Multiple Correlations: Wage Rate Change in 24 Manufacturing Industries, Value Productivity, Concentration, and Unionism [1]

A. Gross value productivity (GVP: sales/production-worker manhour), 1947–59

Equation	Partial correlation coefficients: GVP	Concentration	Unionism	Multiple correlation coefficient
a.	+.352	+.445*		.595
b.	.329		+.720**	.803**
c.	.238	.203	.687**	.812**

B. Net value productivity (NVP: value added/production-worker manhour), 1947–57

Equation	Partial correlation coefficients: NVP	Concentration	Unionism	Multiple correlation coefficient
a.	+.683**	+.095		.728**
b.	.542**		+.578**	.827**
c.	.525**	.019	.572**	.827**

1. See Appendix A, p. 215, for the basic value productivity data and their sources.

*Coefficient is significant with .95 confidence.
**Coefficient is significant with .99 confidence.

said, pay a high price in the credibility of the correlations they generate. The possibility of a reverse correlation (a rise in "productivity" which results from rather than causes a wage rate increase) is much greater than it was with physical productivity. Physical productivity will rise as a result of wage rate increase only if the increase induces a substitution of capital for labor or if there is an "economy of high wages" effect on worker effort or ability to produce. A rate increase will raise sales/manhour, though, any time it increases costs and the burden is shifted to consumers via product prices. Value added/manhour will rise whenever sales/manhour rise and also when (for any other reason) the sum of wages, depreciation, taxes, and profits rise relative to employment. Each successive productivity measure is thus increasingly risky to interpret and the reliability which one attaches to it depends on one's idea of the likelihood of the causal mechanisms which can lead to spurious correlation. Certainly this likelihood cannot be regarded as negligible here, so the preeminence

of the market cannot be regarded as having been definitely reestablished by the value productivity approach to measuring "market forces."

Conclusions from the Analysis So Far

Each of the three major hypotheses of wage determination has now been tested against data of average wage rate change in 24 manufacturing industries, 1948–59, with the following results:

1) The national pattern bargaining hypothesis was found not to hold in its unqualified version (that all industries gave approximately the same absolute wage rate increase as did the key bargainers), but was found to be not inconsistent with the facts in either of two qualified versions: (a) that industries in a politically and economically interrelated "auto-steel orbit" gave essentially identical cents per hour increases or (b) that industries which bargain on an essentially centralized basis did so.

2) Unionism (the percentage of production and related workers in an industry covered by union contracts) was significantly related to wage rate changes and persisted in being so, even after allowance was made for a number of indicators of "market forces."

3) The market forces by themselves were positively but weakly related to wage rate increases (in the case of concentration ratios, profit rates, and productivity change) or hardly related at all (as with employment and output change); only net value productivity, a question-begging measure, was closely related to rate change.

Qualifications and Unresolved Doubts

There are, though, a number of lingering doubts likely to be present in the mind of the reader about the analysis of this section. To some extent, they can be recognized now and met with needed tests, qualifications, and adjustments.

1) The techniques of measurements used here were without doubt very crude. This is especially true of the "market forces" variables. It is probable that the national pattern bargaining and union power hypotheses received generally fair tests, while the market hypothesis was simply too elusive to measure and test adequately. It is quite unlikely that measurement inadequacies account for the impressive showing made by the union power hypothesis. It is possible, however, to think of forces which were not considered here and could conceivably have created a false illusion of union power. One is urban versus rural location, with the need to attract a growing labor force pressing up on urban, unionized wage rates. This explanation seems more appropriate to an earlier historical period than to the present one. Another possibility is that size of plant is somehow related to rate changes through economies of scale and other technical considerations. This seems a better explanation of high wages in unionized industries than of rapid increases. Other possible complications (e.g. employment structure, overtime work, incentive earnings variations) have been minimized here by the use of rates rather than earnings.

2) The choice of a period for study, too, probably had some effect on

the relative importance allotted to the various variables. The alternative choice of the 1948–57 period, for instance (which has the advantage of extending from one cyclical peak to another), would have altered the partial correlation coefficients, but only very slightly. In particular, employment change is no better an explanatory force over this period than over the longer span (it still turns negative when combined with other variables) and in general the factors which loom large in the analysis here are not displaced. Meyers and Bowlby have already shown that the strength of the wage-productivity relation in postwar data of earnings change depends strongly on the time period chosen.[1] However, they found it in no single case to approach in strength the relationship found by Dunlop for 1923–40.

3) Workers in nonmanufacturing and in white collar jobs are becoming increasingly important numerically and a similar analysis of their wage trends would doubtless give different results. Data on their wages are still scarce, though, so such a study is, on any general scale, unfeasible. Speculatively, employment change might have emerged as an important factor in such groups and the relation with unionism and concentration might have been weaker.[2]

4) As has been explained (see above, p.), wage rate changes were used as the dependent variable here out of deference to the traditional view of rates as the "price of labor" of microeconomic theory. The fact that average hourly earnings have been used in most past studies, though, in addition to their considerable claim for theoretical and policy importance, argues for redoing the analysis with average hourly earnings as the dependent variable. This is done below in Section 4.

5) The national pattern bargaining hypothesis has so far appeared to be in accord with the facts, but the notions that unionism per se and (seemingly to a lesser extent) other factors, such as concentration, profits, and productivity change, are important have also survived. The question, then, is just where the pattern fits into this picture of a number of significant influences. How is the existence of three hypotheses, none of which have been disproven, to be reconciled? A departure from the aggregate data and correlation technique and a look at the timing of wage rate changes in a number of specific instances will be needed to clarify further the pattern's role in rate setting.

3. Wage Chronologies and the National Bargaining Pattern

From a study of the aggregate data, one gets the impression of a closely integrated pattern of wage bargaining, whose membership roughly coincides with the "auto-steel orbit" grouping of industries. The intrinsic limitation of the aggregate analysis, though, is that it is impossible to tell whether the essential equality of the cumulated increases of all these units was achieved by

---

1. Fredric Meyers and Roger L. Bowlby, "The Inter-Industry Wage Structure and Productivity," Industrial and Labor Relations Review (October 1956), pp. 26–39.

2. But see pp. 209–11, below.

their consciously following a key bargain or whether the whole thing is coincidental or, more probably, the result of other common influences. And even if conscious pattern bargaining is involved, the question of the independent influences of the pattern remains: do the mere existence of a key bargain and political and market interrelations create strong pressures for the equalization of rate increases or does the impact of coincidental market and power forces simply express itself through the mechanism of imitative wage behavior?

No answer to these questions is possible through the research methods of the preceding section. A disaggregated approach is needed if further progress is to be made. A look at the precise timing of some specific wage changes may help to provide some answer. Testing for the existence of pattern bargaining by looking at the timing of rate changes is not so straightforward as might at first be supposed. The mere fact that one rate change followed another similar one in time does not, of course, prove that it consciously imitated it. On the other hand, in cases of imitation, no one timing pattern need necessarily apply. Depending on the timing of contract expirations, a considerable period of time may go by between the wage rate change of the leader and that of the follower. It can even be argued that followers can anticipate the leader's actions and sign "follower" agreements before the leader signs his. Despite all these difficulties of interpretation, studying timing can be revealing. Avoidance of the post hoc ergo propter hoc fallacy is accomplished by checking for the existence of political and market interrelations to support the pattern bargaining thesis. As to the time lag question, the fact is that the great bulk of imitative agreements are signed with a lag of only a few months behind the signing of the key agreement. In a long-established bargaining pattern, follower agreements are closely aligned with the key bargain in time, being designed to expire shortly after it. If a key bargain has not been struck when the follower agreements expire it is not uncommon (as, for instance, in the steel and related primary metal industries) for negotiations in the follower units to be suspended pending an outcome of the key negotiation.

Three questions will be asked in this section: (1) is there, at any given point in time, a clear-cut national key bargain which potential pattern bargainers could, if they choose, imitate, (2) if so, how extensive is the orbit of imitators of the pattern, and (3) how great is the independent significance of the pattern? The last question (although it may not seem so) can also be answered through a study of the timing of individual rate changes. The best test of the independent force of pattern bargaining is the behavior of units in which there has been, for a time, a coincidence of pattern and other forces on wages and a consequent tendency toward equalization and in which a parting of the ways then occurs. What happens, for instance when a former follower undergoes a marked deterioration in profitability? Does pattern bargaining continue or is it abandoned? The answer to this question can help break the dilemma of too many hypotheses that are in seeming accordance with the facts.

The Key Bargain

The data to be used to study individual wage changes come out of the Wage Chronology Series of the Bureau of Labor Statistics.[1] These chronologics provide detailed accounts of the bargaining activities of a number of the major firms in the economy. The wage rate change portions of several of them are presented in a comparative fashion and in detail in Appendix B.

For there to be a national bargaining pattern or even an auto-steel orbit it is necessary that there be a clear-cut key bargain for the follower units to imitate. When faced with the necessity of specifying a single-valued cumulated national key bargain earlier in this paper (see pp. 183–84, above), the author uncritically followed the opinion of Maher and Soffer that such a bargain actually exists and also their methods of identifying it. The choice was always between the UAW–GM and the U.S. Steel–Steelworkers bargains and it was made on the basis of which bargain in fact seemed, after detailed examination, to be furnishing the model for most followers to imitate. In this question, Soffer's generalization that the unescalated steel contract usually provided the key in periods of reasonable price stability and the auto agreement, with its escalator clause, led in periods of substantial price inflation, was held to offer some guide.[2] But this criterion should be reexamined now. As Table 7 (below) and B-1 (p. 216) show, the two agreements often differed substantially in amount of increase. Did they, however, show at least a general tendency to move together, taking turns at leading, as the national pattern bargaining hypothesis would require?

The third (1948) "round" of postwar collective bargaining was opened by GM's 11 cent increase in May and was followed, after a fashion, by steel and subsequently by many other units. Since the GM agreement featured an escalator clause as well as an immediate increase and steel and most other units did not at once institute wage escalators, the steel rise was put at 13 cents, in anticipation of future consumer price increases. The temporary stabilization of consumer prices in the months that followed left the two agreements close together in amount of cumulated increase.

With the rapid inflation of 1950–51, the difference in form of agreement took on considerable importance. Rates at General Motors rose swiftly and even though steel's anticipatory 15.5-cent increase in 1950 put it ahead briefly, it soon slipped behind again. During these two years the form of the auto contract did indeed make it the key bargainer and steel only another follower. By 1952, however, steel had caught up and the next three years show the two bargains moving in close conformity with each other, creating an unequivocal national key bargain for all to see.

In 1955 the auto workers got their "annual improvement factor" raised from 5 to about 6.1 cents but the steelworkers got 15.2 cents and pushed their cumulated increase 10 cents above the auto industry total. Whether

---

1. See above, p. 181.

2. "The Effects of Cost-of-Living Wage Escalators . . . ," op. cit., p. 33.

this was done in anticipation of consumer price increases which never materialized or not, its effect was to establish permanently a cumulated steel increase about ten cents above the auto increase. From this point on it becomes especially difficult to talk about a single national key bargain. The three-year steel contract signed in 1958 underlined the new dichotomy by instituting an escalator clause which, along with scheduled increases which were higher than those of the auto workers, guaranteed the steel workers at least the maintenance of their advantage vis-à-vis the auto workers. From the two cumulated totals as of December 31, 1959, one might conclude that the gap in favor of the steelworkers has been closing somewhat, but this is strictly a product of the choice of that date as an arbitrary cut-off point. The steel settlement that came at the end of an extended strike in January 1960 put the steelworkers well in the lead again.

The experience of these twelve years shows that the specification of a single national key bargain is very difficult and must be done using some arbitrary rule, such as Soffer's. When would-be follower units looked for a national key bargain to imitate (if, indeed, they did), they often did not receive a clear mandate. When this occurred, what was their reaction? Did they follow Soffer's rule? Or did they instead simply continue to follow whichever of the two rival leaders they habitually followed, as determined by the closeness of their market ties with either steel or autos? The correct answer is definitely the latter, especially as regards the more recent years of the period.

TABLE 7

Some Interim Cumulated Wage Rate Increases:
U.S. Steel and General Motors

Period	U.S. Steel	General Motors
1948-50	28.5¢	24¢
1948-51	28.5	38
1948-53	53.5	53
1948-55	73.2	63.1
1948-59	116.1	108.7

Source: BLS "Wage Chronology Series," Monthly Labor Review, various issues.

The Steel Orbit

Just as the key bargain is really two bargains of somewhat deviant form and amount, so the auto-steel orbit really, upon close inspection, turns out to be two orbits, one clustered around U.S. Steel and the other around General Motors. In Tables 8 (p. 199) and B-2 (pp. 217-18) are listed some representative members of the steel orbit (the other integrated steel producers, not shown in these tables, had chronologies which were almost identical with that of

U.S. Steel). The evidence indicates a rather closely integrated orbit of pattern bargainers, generally following close on the heels of the key bargainer in regard to both timing and amount of wage rate increase. Several breaks in the pattern are discernible, though. All of these appear to represent reactions to distinctive individual circumstances and not attachment to some outside orbit of pattern bargaining.

For example, the 1948 round was participated in at a lower level by the shipbuilders, whose business was undergoing a catastrophic decline connected with reconversion to peacetime production. When the Korean War signaled the rejuvenation of the industry, shipbuilding caught up and the close coherence of the orbit was reestablished. Anaconda copper mining (which enjoyed an exceptional boom in this Korean period) even moved ahead of the key bargain temporarily in amount of cumulated increase. Through the mid-1950's, though, deviations among firms in the steel orbit were quite slight and by the end of the period the pattern was generally intact, much more so than the final totals in Table 8 suggest. The apparent above-pattern position of Alcoa and the West Coast shipbuilders there was very temporary, being subject to elimination after the January, 1960, signing of the new steel contract. At Anaconda, however, falling profits in the last years of the period led to a surrender of pattern bargaining. The surrender was accomplished by a curious method: a three-year contract was signed in 1956 which had immediate and deferred increases comparable to those in the U.S. Steel contract but which failed to provide an escalator and left Anaconda workers far behind after the consumer price rises of 1957-58.

TABLE 8

Some Interim Cumulated Wage Rate Increases in the Steel Orbit

Period	U.S. Steel	Alcoa	Anaconda	Bethlehem shipyards	West coast shipyards
1948-50	28.5¢	26.5¢	27¢	29.5¢	26¢
1948-53	53	55	58	54.5	56
1948-56	82.7	90.75	83	90.5	87
1948-59	116.1	132.25	96.6	116	124

Source: BLS "Wage Chronology Series," Monthly Labor Review, various issues.

The Auto Orbit

Even more internally coherent than the steel orbit was the auto orbit (see Table 9, below, and Table B-3, pp. 218-19). By 1953 the wage actions of the auto manufacturing "Big Three" had been established as identical in form as well as amount and timing. By contrast, the rubber industry bargainers never adopted the long term contract with escalator clause form of contract which became standard in the automobile industry, retaining the

traditional annual negotiations instead, but nevertheless managed to match very closely the amount of the auto increase. In fact, over the twelve-year period as a whole the cumulated increase in the rubber industry was the same as that of the auto industry to the nearest tenth of a cent! This is a truly remarkable example of the extent to which pattern bargaining can be carried. International Harvester serves here to illustrate a firm which had somewhat looser ties to the orbit but still followed the auto bargain quite closely in both amount and form.

TABLE 9

Some Cumulate Interim Wage Rate Increases in the Auto Orbit

Period	General Motors	Ford	Chrysler	Rubber	International Harvester
1948-50	24¢	24¢	24¢	22.5¢	24¢
1948-54	56	56	56	56	54
1948-57	87.3	87.3	87.3	90.7	92
1948-59	108.7	108.7	108.7	108.7	113.4

Source: BLS "Wage Chronology Series," Monthly Labor Review, various issues.

A National Bargaining Pattern?

The actions of the nation's "key bargainers" are thus seen to have a considerable (though not always overriding) influence on the wage rate changes of the bargaining units which are most closely related to them by institutional and market ties. But what of a national, rather than merely orbital, bargaining pattern? How closely are the loosely similar wage rate movements of the steel and auto orbits matched by bargainers elsewhere in the economy? Table 10, below, and the corresponding Appendix Table, B-4 (pp. 220-21), list six bargaining units, as well as repeating, for comparison purposes, the records of the U. S. Steel and General Motors bargains. All six of these units are large and prominent on the national scene, but what they illustrate most clearly is the weakness of the national pattern bargaining hypothesis as an explanation of wage rate changes in American manufacturing. Some instructive points can be garnered from their experiences.

Through 1950 these companies stayed fairly close to the key bargainers in amount of increase. Even this early, though, it was becoming apparent that Sinclair was going to be a consistent upward deviation and International Shoe a clear-cut downward one. During the years of the early 1950's, many variations emerged. North American Aviation and Armour remained fairly close followers of the key bargains, as did, for a time American Viscose. Bituminous coal had, however, already gone off on its independent program of very rapid rate increases. Through the later years of the period these

deviations only became more pronounced. Of the six potential follower companies (i.e. excluding U.S. Steel and General Motors), three and possibly four had deviated from the key bargain by more than "random" amounts by 1959. Differences in market conditions (as with Sinclair, American Viscose, and International Shoe) or in independent union policy (as with bituminous coal) had dominated over the pattern as influences on wage rate change. By contrast, in the case of Armour, pattern bargaining seems to have dominated over spectacular market conditions.

TABLE 10

Some Cumulated Interim Wage Rate Changes: U.S. Steel, General Motors, and Some Illustrative Variant Cases

Period	U.S. Steel	General Motors	American Aviation	Sinclair	Armour	Bituminous Coal	American Viscose	International Shoe
1948–50	28.5¢	24¢	24¢	30.9¢	22¢	24¢	25¢	9¢
1948–53	53	53	57.5	65.4	44.5	72.3	40.5	19.9
1948–56	82.7	75.2	75.5	93.4	80.5	116.5	46	30.9
1948–59	116.1	108.7	109.7	121.9	118.5	155.1	65.5	40.9

Source: BLS "Wage Chronology Series," Monthly Labor Review, various issues.

Conclusions: Pattern Bargaining and Wage Rate Changes in Manufacturing

This detailed study of a number of prominent bargaining units leaves one with two forceful impressions: (1) Pattern bargaining indeed has its role in determining wage rate changes in American manufacturing, but its orbit of influence is much narrower than proponents of the national pattern bargaining hypothesis imply, being mainly limited to the steel and auto clusters (and, of course, to still narrower intraindustry and local groupings, which were not studied here). (2) Through the twelve years studied here there has been a strong tendency toward deterioration in all pattern bargaining. Today the separate steel and auto groupings remain at least fairly coherent internally but the two seem more and more divorced from each other and each orbit of influence is much narrower than it was in the late 1940's.

Both the static limitations and the secular deterioration of pattern bargaining in American manufacturing seem to point to a common moral. Patterns are always most likely to develop where they are strongly reinforced by a uniformity of economic conditions among bargaining units. This uniformity is, of course, to be expected among firms in the same industry or in important purchaser or supplier industries and such relations do in fact exist among members of both the auto and the steel orbits. It also held over a much broader portion of the economy in the immediate post-World War II years.

The economy at that time was dominated by the immense backlog of demand for almost all types of goods; similar boom conditions in most industries produced similar pressures for wage rate increases. Pattern bargaining afforded a cheap and convenient means of resolving these pressures. More recent economic trends have led to a dissimilarity of environment among industries and this is doubtless reflected in the increasing diversity in amount of wage rate changes.

Should a national pattern of wage rate increase be given an analytical place as an independent force in the interpretation of wage rate movements? The answer to this question probably depends on the period of time over which an answer is desired. The uniform single-valued national key bargain as a major influence is probably a thing of the past, but within the relevant orbits individual key bargains will probably continue to retain some short-run influence. Pattern bargaining is a real-world phenomenon—many rate changes are made immediately after and in imitation of those of some accepted leader—but the possibility that a leader-follower relationship will be modified always exists. Anaconda, West Coast shipbuilding, and American Viscose (at various times) served as examples of this and many more examples could have been cited. Moreover, as the preceding section showed, when accumulated rate increases over periods of time as long as twelve years are examined no great discrepancies in amount of increase are found which appear to be attributable to pattern bargaining and cannot be explained as well or better in other terms. When the current economic and political environment of a bargaining unit comes to be at odds with its past history of pattern bargaining, something must give; from the evidence cited here, it is usually the pattern which breaks down. An appealing conclusion with respect to pattern bargaining is that it is, as Reder says,[1] a convenience: when it provides an easy way of reaching a settlement which is acceptable to both sides it is used; when it obviously does not do so it is generally abandoned.

4. Wage Rate Changes and Wage Hypotheses: A Summary of Conclusions

Data of wage rate change in 24 American manufacturing industries, 1948–59, have been used to test three major hypotheses of wage determination. Two methods of analysis have been employed: cross section correlations and detailed study of the timing of rate changes in some specific bargaining units. The three hypotheses fared as follows:

1) The national pattern bargaining hypothesis (that pressure to imitate a key bargain reached by a single key bargaining unit forces nearly uniform wage rate increases) was found to be factually untrue, except in a very qualified form. The qualifications which must be made are that (1) at most times during the period no one clear-cut national key bargain existed, (2) pattern bargaining was generally confined to fairly narrow groups of bargaining units (the steel and auto orbits being the broadest and most important nationally)

1. M. W. Reder, "The Theory of Union Wage Policy," Review of Economics and Statistics (February 1952), p. 37.

and other units followed more independent courses, and (3) even members of bargaining patterns often tended to leave these patterns when changes in their market conditions came into conflict with pattern bargaining. As the last conclusion implies, the independent strength of pattern bargaining as a force in wage determination was judged to be slight, especially over a period of years.

2) The union power hypothesis (taken here to be a simple assertion that unionism has an independent impact on wage rate changes) passed the statistical tests made on it. The percentage of an industry's blue-collar employment covered by union contracts remained significantly correlated with wage rate increase, even after allowance was made for a number of variables representing market forces.

3) The market hypothesis (that supply and demand—or, for practical testing purposes, profits, product market concentration, employment change, output change, and productivity increase—influence wage rate changes) emerged from the test in an equivocal position. From its generally positive (except, in some instances, for output and employment change) but weak relation, one cannot be sure of its actual significance. One's intuitive feeling is that these things are important and the feeling has only failed to be strongly borne out here; it has not been proven false.

5. Changes in Hourly Earnings and the Hypotheses

The ambivalence of the test results on the market hypothesis using data of wage rate change is one reason for being interested in rerunning the tests using average hourly earnings data. There is some reason to believe (see pp. 179–81, above) that hourly earnings changes will be more closely related to market forces than were rate changes. There are other important reasons for wanting to retest with earnings data: (1) comparability of the conclusions of this paper on the union and market impacts with those of previous studies, which used earnings, and (2) the interest inherent in the industry wage level concept which average hourly earnings represent. [1]

An Empirical Comparison of Rate and Earnings Changes

A final, and fundamental, reason for being interested in hourly earnings changes is the practical consideration that earnings and rate changes are often quantitatively different and certainly cannot routinely be assumed to be identical over given periods of time. Table 11 contrasts cumulated rate and earnings changes over the twelve-year study period. It shows the discrepancies, sometimes great, more often relatively small, between the two measures of wage change. The discrepancy for a given industry will be referred to here as "earnings drift."

Earnings drift, as the term is used here, is the difference between the change in average hourly earnings and the average current employment-weighted

1. See pp. 178–81, above, for a fuller discussion of the rationale for interest in earnings.

TABLE 11

Cumulated Average Wage Rate Change, Average Hourly Earnings Change, and Earnings Drift in 24 Manufacturing Industries, 1948–59

(cents per hour)

Industry	Average wage rate change [1]	Change in average hourly earnings [2]	Earnings drift [3]
Flat glass	88.5	152	63.5
Glassware	91.5	101.5	10
Cement	115.5	124	8.5
Basic steel	117.5	158	40.5
Ferrous foundries	100	109	9
Nonferrous metals	113	121.5	8.5
Tin cans	112	139	27
Farm machinery	110.5	116	5.5
Electrical equipment	99.5	94	- 5.5
Automobiles	111	121	10
Aircraft	113	126	13
Shipbuilding	119.5	109	-10.5
Meat packing	117	119	2
Canning	78	68	-10
Grain milling	101	102	1
Baking	105.5	101	- 4.5
Sugar	107.5	93	-14.5
Textiles	52	48	- 4
Apparel	55.5	38	-17.5
Paper	86	100	14
Chemicals	102	117	15
Petroleum refining	111	134	23
Rubber	106.5	105	- 1.5
Shoes	48.5	51	2.5

1. Compiled from announcements of collectively bargained and management-instituted wage rate changes; see above, pp. 181–83.

2. Bureau of Labor Statistics series, December 1947–December 1959.

3. Change in average hourly earnings minus average wage rate change.

change in straight time wage rates for a given industry.[1] It may, of course, be either positive or negative in sign. It occurs (1) when employment structure changes (i.e. when the relative proportions of various skill groups, incentive and time workers, men and women, overtime and straight time workers, day and night shift workers, and whatever other distinctions cause differing rates to be applied, change) and (2) when worker output under incentive payment schemes varies. These might be called the "mechanisms" of earnings drift. A much more important consideration is its causes.

One possibility should be considered at this point: that all the discrepancies between rate and earnings changes in Table 11 are mainly or even entirely reflections of measuring error. Logically it may be true that earnings can move at variance with rates, one might say, but have they in fact done so, to any significant extent? The feeling that measuring error is the major cause of apparent drift is dispelled by a close examination of Table 11. Two observations from that table dispel it: the magnitude of drift and its nonrandomness:

1) While it is clear that drift figures of one or two cents in either direction could easily, given the fallibility of both the rate and the earnings data, be explained by inaccurate measuring and sampling techniques, it is equally obvious that findings such as the 63.5 cents of drift in the flat glass industry, 40.5 cents in basic steel, and −17.5 cents in the apparel industry must surely have some basis in fact. Such extreme values are not, of course, completely impossible ones to be found at the extreme tails of a normal distribution or error, but the clearly observable irregularities in their distribution among industries prove that random error is not the answer.

2) There is an obvious tendency for an identifiable group of industries to have significantly positive drift and another clear-cut group to have less or negative drift. Depending on how one chooses to define it (and making such a choice would beg the question of how drift is caused), the high drift group is generally coincident with the unionized, high profit, concentrated, auto-steel orbit, durable good, etc., industries. Clearly earnings drift is not random, but is systematically related to one or more of these industry characteristics.

Testing the Hypotheses with Earnings Changes

Having established earnings and rate changes over any given period as logically different and over the 1948–59 period specifically as different in fact (but not having explored the reasons why they are different; see below, pp. 209–11), we can proceed to the question of what difference these facts make for the three hypotheses of wage determination. To reiterate, the four possible test outcomes (not all mutually exclusive) are (1) that all earnings-increasing factors tend to be equalized among industries through pattern

1. This definition corresponds to the measured quantity in Table 11. A better definition would leave wage rate changes constantly employment-weighted and leave all structural changes in employment to be reflected in earnings drift. This concept, however, is impossible to measure in the data used here.

bargaining, (2) that union power raises earnings through all possible mechanisms, not just through wage bargaining, (3) that market factors do the same, and (4) that earnings drift is random and that only the "noise" in the equations rises.

The substitution of earnings for rate changes has relatively little impact on the test for the national pattern bargaining hypothesis; what effects it does have, though, make it appear less plausible in its earnings-using version. In its unqualified variant, that earnings increases in all industries tend toward equality, the hypothesis is even less appealing than it was when wage rate increases were used. In the earnings case the interindustry dispersion in amount of wage increase is in fact a bit greater: the standard deviation rises in absolute terms from 20.8 cents to 30.3 cents and as a percentage of the mean increase ($.984 and $1.061 respectively) from 21 to 29 per cent. In its qualified auto-steel orbit version the hypothesis is again less well in accordance with the facts when earnings are used. Increases within the orbit continue to be significantly different from (i.e. higher than) increase elsewhere. Variance analysis bears this out.[1] However, coherence with the orbit is considerably diminished by the shift from rates to earnings. The standard deviation from the mean increase within the orbit rises from 9.1 cents (or 8.4 per cent of the mean rate increase of $1.083) to 19.8 cents (16.1 per cent of the orbital mean earnings rise of $1.228). Proponents of pattern bargaining hypotheses, it would appear, would be well advised to consider wage rate changes rather than all earnings-increasing contract provisions as the basis for imitative bargaining; even if it is true that pattern bargaining brings about the latter kind of equalization, it seems that interunit differences in experience between contract negotiations create considerable diversity in actual earnings movements.

Unionism's impact remains clearly evident in these earnings data. As a simple rank correlation coefficient unionism assumes a value of +.750, which is statistically significant with .99 probability (the comparable figure obtained using rates was +.737; see p. 186, above). As a look at Tables 12 and 13 (below) shows, it retains its explanatory significance after allowance is made for the other variables. Equation b of Table 13 shows the union power hypothesis passing its severest test; even allowing for productivity change, profitability, market concentration, and the fact of a wide range of earnings-influencing mechanisms in addition to wage bargaining, unionism retains .95 significance.

The most noticeable change in test results produced by the shift from the rate to the earnings variable is found in the relation of wages to profits and industrial concentration, two of the "market forces" measures. Both of these loom as highly significant as simple coefficients and barely miss being significant at the .95 level when allowance is made for differing original wage levels (see Table 12). The other "market forces" variables—productivity change, employment change, and change in output—continue to be positively but weakly correlated with wage change. In multiple correlations which

1. $F = 7.96$; $F_{.99} = 7.88$.

combine the economic and unionism variables (see Table 13; these equations correspond to equations g and h of Table 5, p. 192, above) the seeming significance of market concentration vanishes, but profit rates remain significant or nearly so.

TABLE 12

Simple Correlations: Average Hourly Earnings Changes in 24 Manufacturing Industries versus Unionism and Various Measures of "Market Forces," 1948–59

Independent variable	Rank correlation coefficient	Normal correlation coefficient	As a partial coefficient with the initial wage level
Productivity [1]	.134	.132	.027
Profits	.508*	.560*	.399 [2]
Employment change	.116	.165	.234
Output change	.208	.188	.088
Concentration	.589**	.608**	.393 [2]
Unionism	.750**	.752**	.550**

1. The variables are defined as before (in Table 2, p.); for precise definitions and a discussion of their reliability, see Appendix A.

2. Coefficient barely missed being significant at the .95 level ($r_{.95} = .406$)

*Coefficient is significant with .95 confidence.
**Coefficient is significant with .99 confidence.

Another point to be garnered from these correlations is that equations using a given set of variables tend to produce a higher proportion of explained variation when earnings are used as the wage measure than when rates are the dependent variable. The final formulation with earnings succeeds in "explaining" just under three-fourths of the total variation, whereas with rates only a little less than two-thirds was accounted for. This, of course, suggests an orderly pattern in earnings drift and belies the "noise" hypothesis of its causation.

In summary, measuring wage changes in terms of gross average hourly earnings instead of wage rates has the following effects on the hypotheses:

1) The hypothesis of equalization of absolute earnings increases through national pattern bargaining is less well in accord with the facts than the corresponding hypothesis of rate change equalization.

2) Union power appears as clearly to be an influence in earnings changes as in wage rate changes.

3) Of the "market forces," profits and, to a lesser extent, market

TABLE 13

Coefficients of Multiple and Partial Correlation: Average Hourly Earnings Changes in 24 Manufacturing Industries and Various Combinations of the Independent Variables, 1948-59[1]

Equation	Partial correlation coefficients				Multiple correlation coefficient
	Productivity	Profits	Concentration	Unionism	
a.	+.000	+.442*	+.173	+.585**	.833**
b.[2]	-.132	.392	.125	.526*	.854**

1. These equations correspond to equations numbers g and h of Table 5, p. 192, above, except that they substitute earnings for rate changes.

2. Equation b includes as an additional variable the initial average wage rate. It possesses a partial correlation coefficient of about +.347, increases the over-all explanatory value of the equation as shown, and pushes the partial correlation coefficient of profits somewhat below the .95 level of significance.

*Coefficient is significant with .95 confidence.
**Coefficient is significant with .99 confidence.

concentration were much more closely related to earnings than to rate changes.

6. Final Conclusions and Broader Implications

A weak but inevitable conclusion on the three hypotheses of wage determination that this paper set out to test is that each has its role to play. The nature of this role has to some extent been defined here. The impact of pattern bargaining has been found to be more apparent than real and its real impact more short run than long run. The existence of unionism has been found, in this time and place, to appear to make a difference. Various economic measures have shown some relation to wage changes, but only a weak one. What is the broader meaning of these findings? To what extent can they be used as a basis for generalizations on wage determination in general?

In regard to the national pattern bargaining hypothesis, the conclusions of this paper were in line with the conclusions of past studies of narrower bargaining patterns.[1] The study of Seltzer of the basic steel and related industries and that of Levinson of Detroit-area automotive and related firms

1. George Seltzer, "Pattern Bargaining and the United Steelworkers," Journal of Political Economy (August 1951), pp. 319-31; Harold M. Levinson, "Pattern Bargaining: A Case Study of the Automobile Workers," Quarterly Journal of Economics (May 1960), pp. 296-317.

both reached similar conclusions. They found the apparent uniformity of wage and other collective bargaining behavior to be subject to modification in units which were less closely related by market ties to the key bargainer [1] and also to be subject to almost complete abandonment when the high level of key bargain rate increase and the poor financial status of a given firm combined to raise the key bargain rate increase to a "crisis level" from the point of view of that firm. [2] The common conclusion of this paper and those of Seltzer and Levinson corroborate each other and thus seem capable of wide generalization, to all forms of pattern bargaining.

The positive conclusion here on the union power impact is obviously one of much less generality. It accords with Ozanne's similar conclusion, which was based on a nearly coincidental time period but very different research methodology. [3] That it apparently fails to coincide with the conclusions of Bowen and Levinson for similar postwar periods is, as has been seen, capable of explanation on grounds of relatively minor procedural differences. [4] As far as the postwar years are concerned, Ozanne's prediction about the accumulating evidence demonstrating that collective bargaining is more than a ritual seems to borne out. Limitations of scope and time inhibit generalizations from this conclusion however. As already mentioned, the inclusion of white collar and nonmanufacturing segments of the economy in the tests might have weakened unionism's role. That it would necessarily have done so, though, is by no means clear. As Table 14 (below) shows, average hourly earnings changes over the 1948–59 period were greater in absolute terms in some of the excluded unionized sectors than in at least some important nonunionized areas and were also perhaps a bit higher when measured in percentage terms. Further evidence is clearly needed. The limitation in time is certainly a crucial one. As Levinson has demonstrated, [5] unionism appears to have an impact in some historical periods and not in others. Levinson's supposition on the reasons for this variation was that unions were more potent when the over-all level of unemployment was high and when government policy was favorable. Whether, in the historical context, these conditions are particularly characteristic of the 1948–59 period is problematical; a reconsideration of Levinson's broad generalization may be called for.

Finally, there are the ambivalent test results on the market hypothesis. How much importance should be attached to the weak showing of the "market

---

1. ". . . the key bargain in basic steel . . . is most effective in the integrated units, where it serves almost as a rule, and lease effective in the nonintegrated units, where it serves as a goal." Seltzer, op. cit., p. 322.

2. Levinson, "Pattern Bargaining . . . ," op. cit., pp. 306ff.

3. See above, pp. 185–86.

4. See pp. 186–87, above.

5. Harold M. Levinson, Unionism, Wage Trends, and Income Distribution, 1914–1947 (Ann Arbor, Mich.: University of Michigan Press, 1951).

TABLE 14

Changes in Average Hourly Earnings in Some Heavily and Lightly Unionized Industries Which Were Excluded from the Study, 1948–59

Industry	Absolute earnings increase [1]	Percentage earnings increase [1]
Heavily unionized industries		
Mining	$1.03	68.4%
Contract construction	1.38	84.7
Class I railroads	1.28	99.2
Lightly unionized industries		
Wholesale trade	.95	75.4
Retail trade	.64	69.6

1. These are increase from December 1947 to December 1959.

Source: Bureau of Labor Statistics series.

forces" variables here? As has already been suggested, the answer is probably, "very little." The measurement problem and the difficulty of putting the competitive theory of wage determination into testable form were just too great for a valid test to be obtained. One interesting point, though, is the improved showing made by the market variables, especially the rate of profit on stockholders' equity, when earnings changes rather than rate changes are considered. This, of course, helps explain the fact that higher correlations were obtained in past studies of factors related to wage change (which used average hourly earnings as the wage measure) than in Section 2 of this paper, where rates were used. Of more basic interest, though, is the question of why different results occurred when earnings were substituted for rates. This, of course, hinges on the question of why earnings changes were in fact of a different magnitude from rate changes. Some of the "mechanisms" of earnings drift were mentioned above (pp. 203–05) and also one major hypothesis about its fundamental causation (see pp. 179–81). This hypothesis, a product of studies of "wage drift" in wartime and "wage restraint" economies, had earnings drft bringing earnings into closer accord with a market-determined equilibrium than rates, which were thought to be set by government, employer, or union policy (or some combination of these), with scant regard for market considerations. Has earnings drift served some such function in the present case? The higher correlations with market forces of earnings than of rate changes suggest that it might have done so. Or is some different

line of causation involved, perhaps that suggested recently by Dunlop?[1] On the other hand, the complementarity, already noted, of earnings drift and wage rate increases, could be taken to imply that rate and earnings increases are, by and large, produced by the same causal forces. Only detailed empirical study can delineate the causation of earnings drift; it is left in the form of an open question here.

1. "Wage determination most fruitfully ought to be divided into two parts — the dtermination of the wage rate schedule and the determination of the wage drift, defined as the change in hourly earnings resulting from other than changes in scheduled wage rates. Changes in the wage drift are a function of levels of output, unemployment, rates of technical change and of premium hours. Changes in scheduled wage rates are primarily a function of profits and consumers prices." John T. Dunlop, "Policy Problems: Choices and Proposals," in Charles A. Myers (ed.), Wages, Prices, Profits and Productivity, New York (American Assembly), 1959, p. 149.

Appendix A: Sources and Evaluation of Independent Variables Used in Sections 2 and 5

TABLE A-1

Productivity Change, Profit Rate, Employment Change, Concentration Ratio, Unionization, and Initial Average Wage in 24 Manufacturing Industries, 1948-59

Industry	Productivity Change [1]	Profit Rate [2]	Employment Change [3]	Concentration Ratio [4]	Degree of Unionization [5]	Initial Average Wage [6]
Flat glass	51.0%	15.5%	3.2%	99%	5	$1.50
Glassware	17.0	14.3	5.7	64	5	1.14
Cement	67.5	14.8	4.2	31	5	1.18
Basic steel	41.1	10.6	- 20.2	67	5	1.49
Ferrous foundries	24.9	7.1	- 17.3	25	5	1.32
Nonferrous metals	53.7	11.4	- 11.6	65	5	1.34
Tin cans	28.6	10.3	17.4	80	5	1.20
Farm machinery	70.8	10.2	- 19.8	52	5	1.38
Electrical equip.	73.4	12.7	19.0	49	4	1.28
Automobiles	40.2	15.9	- 11.9	91	5	1.50
Aircraft	122.4	14.4	157.7	48	5	1.37
Shipbuilding	56.3	7.3	- 16.5	40	5	1.48
Meat packing	31.1	5.4	7.3	35	5	1.17
Canning	53.3	8.1	- 10.2	31	4	1.07
Grain milling	37.2	11.8	- 21.7	33	3	1.10
Baking	25.2	12.2	- 14.2	29	3	1.09
Sugar	98.4	6.6	- 25.1	65	5	1.04
Textiles	84.6	6.2	- 30.3	27	2	1.09
Apparel	22.7	5.1	4.8	14	3	1.11
Paper	58.4	11.2	10.4	23	4	1.16
Chemicals	101.0	13.0	1.2	43	4	1.23
Petroleum refining	79.6	13.9	- 16.5	32	5	1.53
Rubber	56.5	11.7	- 9.4	55	5	1.39
Shoes	19.3	9.1	- 5.1	29	3	1.04

1. Change in output per production-worker manhour, calculated using the Federal Reserve production index and the relevant BLS figure on manhours of employment of production and related workers (i.e. employment times average weekly hours). The procedure has its shortcomings, since some of the components of the production index were themselves originally estimated from data on labor input, rather than from actual measures of physical output. It is, though, the only way in which comparable productivity data can be compiled for all these industries. Better productivity indexes

have been compiled only on a selective basis. Where the indexes used here could be compared with the superior indexes, though, the discrepancies observed were very slight.

2. Average after-tax profits as a percentage rate of return on stockholders' equity. Sources: United States Federal Trade Commission–Securities and Exchange Commission, Quarterly Financial Report on Manufacturing Corporations; Moody's Industrial Manual.

3. Percentage change, 1959 over 1947, of manhours of employment of production and related workers.

4. Percentage of shipments made by the four largest corporations in 1954 (in the case of the three-digit industries). Figures for two-digit industries are not strictly comparable, being weighted averages of the concentration ratios in the component three-digit industries. This is not identical with percentage of shipments by the four largest firms in the two-digit industry itself. The approximation was judged close enough for present purposes, though. Source: United States Senate, Committee on the Judiciary, Concentration in American Industry, 1957.

5. Percentage of all production and related workers covered by union contract. For two-digit industries the source is H. M. Douty, "Collective Bargaining Coverage in Factory Employment, 1958," Monthly Labor Review (April 1960), pp. 345–49. For most three-digit industries, the source is an earlier BLS study which pertains to 1946 (Extent of Collective Bargaining and Union Recognition, 1946, Bureau of Labor Statistics Bulletin 909, Washington, 1947); these earlier figures have been checked and adjusted for approximate correspondence with the 1958 estimates for the broader two-digit industries which they go to make up. For two industries (ferrous foundries and tin cans), estimates were made by the present writer on the basis of the coverage obtained of industry employment in announcements of collectively bargained wage settlements. The diversity of source and probable roughness of the estimates make it advisable to consider industries as being split only into broad groups with respect to unionism.

I. 0–20% under contract	III. 41–60	V. 81–100
II. 21–40	IV. 61–80	

The use of these broad categories in lieu of precise figures enhances slightly the correlation coefficients attached to unionism. The cluster of unionization figures around a few values with little spread between them means that for a given correlation the sum of the deviations of the expected from the actual values is likely to be about the same as if the precise figure were used, but the sum of the deviations squared is likely to be somewhat smaller. The correlation is thus slightly better. Careful checking indicates, however, that the impact of this effect is slight. Take the rank correlation of unionism with average wage rate change, +.775, as an example. Had the riskier attempt been made to specify ten groupings (deciles) rather than only five, the resulting

coefficient would have fallen to +.723 (still clearly significant) because of the great reduction in the number of tied ranks. Had the investigation been further limited to the 14 industries for which the proper decile classification was most certainly known, the fall of the coefficient would have been only to +.763. Thus, while the prevalence of tied ranks creates some upward bias in all the coefficients attached to unionism, it is clear from these illustrative cases that the magnitude of the bias is not great. By no stretch of the imagination could the significance of unionism in these tests be ascribed to it.

6. Level of straight time average hourly earnings (in fact, the BLS gross hourly earnings series corrected by standard BLS methods for the probable impact of overtime) in December 1947. Straight time earnings are appropriately used to represent the average basic wage rate at the start of the period because at any point in time, using current employment weights, the two concepts are identical. They become at variance with each other only as employment weights change and incentive earnings vary through time.

TABLE A-2

Change in Gross and Net Value Productivity in 24 Manufacturing Industries

Industry	Change in GVP, 1948-59[1]	Change in GVP, 1948-57[1]	Change in NVP, 1948-57[2]
Flat glass	133.4%	109.3%	59.2%
Glassware	165.0	114.4	88.5
Cement	253.6	147.2	144.5
Basic steel	102.1	66.9	164.6
Ferrous foundries	141.8	86.7	75.7
Nonferrous metals	182.9	121.7	109.8
Tin cans	202.5	184.3	131.0
Farm machinery	184.8	163.1	131.6
Electrical equipment	214.2	149.1	98.6
Automobiles	164.3	144.4	125.1
Aircraft	314.8	244.1	155.9
Shipbuilding	180.2	139.7	82.7
Meat packing	-2.9	-8.6	80.5
Canning	196.1	176.7	82.7
Grain milling	118.4	96.3	79.0
Baking	124.6	98.0	103.4
Sugar	173.6	149.6	96.2
Textiles	112.1	99.7	34.0
Apparel	163.9	131.8	30.6
Paper	99.5	73.8	75.7
Chemicals	146.3	108.7	118.5
Petroleum refining	178.1	152.3	92.8
Rubber	165.0	94.4	88.0
Shoes	100.7	77.8	48.5

1. Net sales per production-worker manhour. Sources: Survey of Business, various numbers, for sales; BLS for employment and hours.

2. Value added per production-worker manhour. Sources: United States Bureau of the Census, Census of Manufactures, Annual Survey of Manufactures, various editions, for value added; BLS for employment and hours.

Appendix B: Chronologies of Wage Rate Change, 1948-59

The following tables present comparative chronologies of wage rate change, respectively of the key bargainers, the steel orbit, the auto orbit, and an assorted group of illustrative cases. The data source in each case is the "Wage Chronology Series" of the Bureau of Labor Statistics, published at irregular intervals in the Monthly Labor Review. Tables B-1 to B-4 correspond with the tables of cumulated total rate increases, Tables 7 to 10, in the text above.

These tables can be read as follows. Rate changes are grouped into three categories: (1) newly bargained increases, which are underlined, (2) previously bargained (deferred) increases, which are marked with no special notation, and (3) formula (escalator) changes, which are enclosed in parentheses. They are all listed by the year and quarter in which they became effective. Pattern bargaining appears most evidently in these tables as a lining up across the page of similar negotiated (i.e. underlined) increases. An example of clear pattern bargaining is the "third round" of postwar rate increases in the second and third quarters of 1948. For further elucidation of the tables, see the text and Tables 7-10, above (pp. 197-201).

TABLE B-1

The Timing of Wage Rate Changes at U.S. Steel and General Motors, 1948-59

	U.S. Steel	General Motors		U.S. Steel	General Motors
1948 I			1954 I		(-1)
II		11	II		5, (-1)
III	13	(3)	III	5	(1)
IV			IV		(-1)
1949 I		(-2)	1955 I		
II		3, (-1)	II		6.1
III			III	15.2	(1)
IV			IV		
1950 I		(-2)	1956 I		(-1)
II		4	II		6.1, (1)
III		(5)	III	9.5	(4)
IV	15.5	(3)	IV		(2)
1951 I		(5)	1957 I	(3)	(1)
II		4, (3)	II		6.1, (2)
III		(1)	III	8.2, (4)	(3)
IV		(1)	IV		
1952 I	16	(3)	1958 I	(5)	(3)
II		4, (-1)	II		
III		(3)	III	8.2, (4)	6.2, (2), (1)
IV		(-1)	IV		
1953 I			1959 I	(1)	
II	8.5	5, (-1)	II		
III		(1)	III		6.2, (2)
IV		(2)	IV		(1)

TABLE B-2

The Timing of Wage Rate Changes at U.S. Steel and in the "Steel Orbit," 1948-59

		U.S. Steel	Alcoa [1]	Anaconda [2]	Bethlehem shipyards	West Coast shipyards
1948	I					
	II		12			
	III	13		12	7	8
	IV					
1949	I					
	II					
	III					
	IV					
1950	I			5		
	II					
	III		14.5			6
	IV	15.5		10		
1951	I				22.5	12
	II					
	III			15.5		12
	IV					
1952	I	16	15			
	II				18	
	III		4	15.5		15
	IV					
1953	I		1			
	II	8.5			7	
	III		8.5			3
	IV					
1954	I		1			
	II					
	III	5	5	2	3	6
	IV				2	
1955	I					
	II					
	III	15.2	15	13	15	7
	IV					
1956	I					
	II		3			
	III	9.5	11.75	10		18
	IV				16	
1957	I	(3)				
	II					
	III	8.2, (4)	10.5, (4)	6.8	8.5	15
	IV					

Table B-2 continued.

	U.S. Steel	Alcoa [1]	Anaconda [2]	Bethlehem shipyards	West Coast shipyards
1958 I	(5)	(5)			
II		(4)			
III	8.2, (4)	8	6.8	9, (7)	11
IV					
1959 I	(1)	(1)		(1)	
II					
III					10
IV					

1. The precise dating employed here refers to the contract with the United Steelworkers; the company's agreement with the Aluminum Workers has had a similar but not identical history.

2. Applies only to mining operations and to the Mine, Mill, and Smelter Workers contract.

TABLE B-3

The Timing of Wage Rate Changes at General Motors and in the "Auto Orbit," 1958-59

	General Motors	Ford	Chrysler	Rubber [1]	International Harvester [2]
1948 I					
II	11			11	11
III	(3)	13	13		
IV					
1949 I	(-2)				
II	3, (-1)				
III					
IV					
1950 I	(-2)				
II	4				
III	(5)	8	10	11.5	10
IV	(3)	(3)	(1)		(3)
1951 I	(5)	(5)	(5)		(5)
II	4, (3)	4, (3)	4, (3)		(3)
III	(1)	(1)	(1)	12	4, (1)
IV	(1)	(1)	(1)		(1)
1952 I	(3)	(3)	(3)		(3)
II	4, (-1)	4, (-1)	4, (-1)		(-1)
III	(3)	(3)	(3)	10	4, (3)
IV	(-1)	(-1)	(-1)		(-1)

Table B-3 continued

	General Motors	Ford	Chrysler	Rubber [1]	International Harvester [2]
1953 I					
II	5, (-1)	5, (-1)	5, (-1)	5	(-1)
III	(1)	(1)	(1)	5	4, (1)
IV	(2)	(2)	(2)		(1)
1954 I	(-1)	(-1)	(-1)		
II	5, (-1)	5, (-1)	5, (-1)		(-1)
III	(1)	(1)	(1)	6.5	4, (1)
IV	(-1)	(-1)	(-1)		(-1)
1955 I					
II	6.1	6.1			
III	(1)	(1)	6.1, (1)	14	13, (1)
IV					
1956 I	(-1)	(-1)	(-1)		(-1)
II	6.1, (1)	6.1, (1)	6.1, (1)		(1)
III	(4)	(4)	(4)	6.2	6, (4)
IV	(2)	(2)	(2)		(2)
1957 I	(1)	(1)	(1)		(1)
II	6.1, (2)	6.1, (2)	6.1, (2)		(2)
III	(3)	(3)	(3)	14.5	6, (3)
IV					
1958 I	(3)	(3)	(3)		(3)
II					
III	6.2, (2), (1)	6.2, (2), (1)	6.2, (2), (1)	8	(1)
IV					
1959 I					6.1
II					
III	6.2, (2)	6.2, (2)	6.2, (2)	10	6.3, (2)
IV	(1)	(1)	(1)		(1)

1. The numbers refer to U.S. Rubber, but the other companies of the "Big Four" have had an almost identical history.

2. Applies to the United Auto Workers contract, but that of the Farm Equipment Workers (later absorbed into the United Electrical Workers) is similar.

TABLE B-4

The Timing of Wage Rate Changes at U.S. Steel, General Motors, and in Some Illustrative Variant Cases, 1948–49

	U.S. Steel	General Motors	North American Aviation	Sinclair	Armour[1]	Bituminous Coal	American Viscose	International Shoe
1948 I								(3)
II		11			9		15	(-3)
III	13	(3)	10	17.5		14.3		(3)
IV								(3)
1949 I		(-2)						(-3)
II		3, (-1)						
III			5					
IV					2			
1950 I		(-2)						
II		4				9.7		
III		(5)	9		11		10	
IV	15.5	(3)		13.4				6
1951 I		(5)	(2)		11.5	22.1	3	
II		4, (3)	(7)	7.5				
III		(1)	(1)				3	
IV		(1)	(1)	(1)	6		5.5	6
1952 I	16	(3)	(3)	(3)				
II		4, (-1)	10, (-1)	15, (-1)				
III		(3)	(2)					4.5
IV		(-1)	(1)		5	26.2	4	
1953 I								
II	8.5	5, (-1)	(-3)					
III		(1)		9	5			
IV		(2)	8.5, (2)					
1954 I		(-1)						
II		5, (-1)						
III	5	(1)			5			
IV		(-1)	5					
1955 I				10				
II		6.1					5.5	
III	15.2	(2)			14	16.6		7
IV								
1956 I		(-1)	10	16				4
II		6.2, (1)				11		
III	9.5	(4)	(1)					
IV		(2)	(2)		12	16.6		
1957 I	(3)	(1)	7, (2)		(2)			
II		6.1, (2)	(2)	15		11	7	
III	8.2, (4)	(3)	(2)		8, (3)			
IV			(3)					

Table B-4 continued

	U.S. Steel	General Motors	North American Aviation	Sinclair	Armour [1]	Bituminous Coal	American Viscose	International Shoe
1958 I	(5)	(3)	(1)		(4)			
II			3.7, (2)				0.5	
III	8.2, (4)	6.2, (2), (1)	(2)		7.5, (4)			
IV								6
1959 I	(1)		(1)	13.5	(1)	16.6		
II			7.5, (-1)			11	12	
III		6.2, (2)	(1)		8.5			
IV		(1)	(1)					4

1. The dating applies to contracts with the United Packinghouse Workers: those of the Meat Cutters and Butcher Workmen and the National Brotherhood of Packinghouse Workers have been similar.

Summary: This essay is a shortened version of a longer dissertation submitted to the Economics Department of Yale University. The purpose of this dissertation was the investigation of the relationship between the major oil refiners and their service station operators. A legal study, omitted from this essay, discussed cases in which antitrust law has been applied to exclusive dealing, the refiners' requirement that their dealers dispense certain brands of petroleum and nonpetroleum products to the exclusion of other brands. A discussion of the existing relationships, summarized in this essay, between the refiners and their dealers, concludes that despite the legal proscription of this practice exclusive dealing persists in these relationships. These relationships are consistent with four complementary explanations. First, the use of lessee dealers permits the refiners to minimize their distributional costs. Second, refiners prefer some property control over their station outlets to guarantee access to final demand for their products. Third, sponsorship of particular tire, battery, and accessory brands and the resulting exclusion of alternative brands allows the refiners to gain substantial revenue from granting privileged access to these stations to the favored TBA suppliers. Fourth, exclusive dealing permits the refiners to maximize their rental returns from leased stations. The final section presents data which are consistent with this last explanation.

This essay was submitted to Yale University in 1962 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics at Wesleyan University.

Dissertation Committee: Dean John Perry Miller (Chairman), Professors Charles H. Berry and Kent T. Healy.

*

I wish to express my appreciation for the help and encouragement of Dean John Perry Miller and Professors Charles H. Berry, Kent T. Healy, and Ward S. Bowman, all of whom have contributed substantially to this study at its various stages. I am grateful for the financial assistance made available by a Ford Foundation Doctoral Dissertation Fellowship and by a Faculty Research Grant from the Trustees of Wesleyan University. Members of the petroleum industry have generously devoted many resources to gathering and providing me with information in many forms; by their choice they remain anonymous.

R. A. M.

EXCLUSIVE DEALING IN THE PETROLEUM INDUSTRY: THE REFINER-LESSEE DEALER RELATIONSHIP

Richard A. Miller

1. Introduction

Major oil companies have often urged their service station dealers, with varying degrees of insistence, to handle certain product brands rather than other brands. In the extreme, major oil companies have conditioned the sale of their gasoline or the lease of their stations on the exclusion of brands of petroleum products, tires, batteries, and accessories which they did not sell or sponsor.

Section Three of the Clayton Act proscribes these exclusive dealing requirements "where the effect . . . may be to substantially lessen competition or tend to create a monopoly in any line of commerce."[1] Applying this section to the relationship between one major and its lessee station operators, the U.S. Supreme Court struck down the exclusive dealing employed by the Standard Oil Company of California.[2] While admitting that it practiced exclusive dealing, Standard denied that its contracts with its dealers satisfied the qualifying clause of Section Three. Speaking for a bare majority of a divided Court, Mr. Justice Frankfurter set the fundamental issue thus:

> The issue before us, therefore, is whether the requirement of showing that the effect of the agreements "may be to substantially lessen competition" may be met simply by proof that a substantial portion of commerce is affected or whether it must also be demonstrated that competitive activity has actually diminished or probably will diminish.[3]

Although "requirements contracts . . . may well be of economic advantage to buyers as well as to sellers, and thus indirectly of advantage to the consuming public,"[4] the Court is "ill-suited" to engage in the necessary economic analysis for such determination:

> We conclude, therefore, that the qualifying clause of §3 is satisfied by proof that competition has been foreclosed in a substantial share of the line of commerce affected. It cannot be gainsaid that observance by a dealer of his requirements contract with Standard does effectively foreclose whatever opportunity there might be for competing suppliers to attract his patronage, and it is clear that

---

1. 15 U.S.C. §14 (1958). The thesis of which this essay is a shortened version is available to interested readers at Sterling Memorial Library, Yale University.

2. Standard Oil Company of California v. United States, 337 U.S. 293 (1949). This decision is often referred to as "Standard Stations."

3. 337 U.S. 293, 299.

4. 337 U.S. 293, 306.

> the affected proportion of retail sales of petroleum products is substantial. In view of the widespread adoption of such contracts by Standard's competitors and the availability of alternative ways of obtaining an assured market, evidence that competitive activity has not actually declined is inconclusive. Standard's use of the contracts creates just such a potential clog on competition as it was the purpose of §3 to remove, wherever, were it to become actual, it would impede a substantial amount of competitive activity.[1]

This Standard Stations decision has been the subject of various interpretations.[2] For example, it is not at all evident which "share of the line of commerce" the Court considered relevant. Varied market shares falling to Standard are mentioned in the Court's opinion.[3] Of the several product sales in the seven state "Western area" Standard accounted for

> 2% of the tire sales through its branded stations leased to dealers
> 2% of battery sales through its branded stations leased to dealers
> 5% of lubricating oil through its branded stations leased to dealers
> 6.7% of gasoline sales through its branded stations leased to dealers
> 13.5% of gasoline sales through its branded stations leased to dealers and managed by employees
> 23% of all gasoline sales, including sales to commercial accounts

> and along with the six other major refiners in this area, which also used exclusive dealing contracts with their service station dealers,
> 56% of all gasoline sales through service stations.

Also, Standard controlled through ownership or lease 16% of all station outlets in this market area. The relevance of the International Salt decision—the only Section Three precedent that the Court found relevant—implies that the quantity of commerce rather than the share of the market was the determining factor in the Standard Stations decision: no market share was mentioned in the Supreme Court's International Salt opinion.[4] And of what importance was the collective but not collusive[5] use of exclusive dealing by all the major oil companies

---

1. 337 U.S. 293, 314.

2. Derek C. Bok, "The Tampa Electric Case and the Problem of Exclusive Arrangements Under the Clayton Act," The Supreme Court Review, 1961, ed. Philip B. Kurland (Chicago: The University of Chicago Press, 1961), pp. 267–332, esp. 275–77, summarizes some of these interpretations.

3. 337 U.S. 293, 295–97.

4. International Salt Co. v. United States, 332 U.S. 392 (1947) International Salt dispensed under tying agreements about half a million dollars of salt products annually, a not insubstantial amount. Standard's gasoline sales annually were $57 million, sales of other products, $8 million.

5. 337 U.S. 293, 309.

selling in Standard's market area? Another possibility, mentioned by Frankfurter in his Motion Picture Advertising dissent, is the "discrepancy in bargaining power" between the refiner and the dealers.[1]

Recently the Supreme Court has dispelled some of these difficulties of interpretation, but many remain. In the Tampa Electric Case, the emphasis is placed on the substantiality of the market share:

> . . . the competition foreclosed by the [exclusive dealing] contract must be found to constitute a substantial share of the relevant market.[2]

Citing the Standard Stations decision, the Court went on to mention the

> large number of gasoline stations—5,937 or 16% of the retail outlets in the relevant market—and the large number of contracts, over 8,000, together with the great volume of products involved.

In studying the substantiality and probable effect of the exclusive requirements, other factors must also be considered:

> . . . the relative strength of the parties, the proportionate volume of commerce involved in relation to the total volume of commerce in the relevant market area, and the probable immediate and future effects which the pre-emption of that share of the market might have on effective competition therein. It follows that a mere showing that the contract itself involves a substantial number of dollars is ordinarily of little consequence.[3]

And, of course, what divides the substantial and not substantial lessening in terms of market percentage?

It is not difficult to sympathize with the Supreme Court's reluctance to undertake complete analysis of each exclusive dealing case. Easily applied rules benefit the courts as well as businessmen in their interpretations of socially acceptable behavior. On the other hand, the qualifying clause of Section Three does imply questions of economic analysis. As Derek C. Bok has mentioned,

> Like so many phenomina in the world of antitrust law, exclusive arrangements are only partially understood. Little empirical work has been done on their nature and effects, and almost no effort has been made to devise a theoretical framework in which they can be understood and evaluated.[4]

---

1. Federal Trade Commission v. Motion Picture Advertising Service Co., 344 U.S. 392, 402 (1953).

2. Tampa Electric Company v. Nashville Coal Company, 365 U.S. 320, 328 (1961).

3. 365 U.S. 320, 329.

4. "The Tampa Electric Case . . .," p.271.

This essay probes the exclusive arrangements between the major oil refiners and their service station operators in an attempt to contribute to the understanding of these arrangements. The following sections summarize the existing refiner-dealer relationships and suggest several interpretations for them.

2. Relationships between Refiners and Their Dealers

The discrepancy in size, when size is measured by assets or sales, between major oil companies and any one of their service stations is enormous. The majors, comprising the largest twenty or so petroleum refiners, are generally vertically integrated from the mining of crude oil through refining and transportation to distribution to final customers. The smallest of these twenty majors, the Richfield Oil Corporation, in 1960 had assets of $289 million, while the assets of the largest, the Standard Oil Company (New Jersey), were $8.7 billion. Gross sales for each of the twenty majors range from several hundred million dollars annually to several billion dollars. Fourteen of these twenty firms distribute their products in twenty or more states each, and together the twenty produce more than eighty percent of the gasoline sold in this country.

In contrast the service stations are numerous and small; the 1958 Census of Business classified over 206,000 establishments as service stations [1] with aggregate gross receipts of over $14 billion, or about $70,000 each on the average. Over eighty percent of the stations are individual proprietorships, and thirteen percent are partnerships. Together these account for 85% of the sales through all stations. The remaining seven percent of the stations, predominately corporations, sell almost fifteen per cent of the products.

Between the refiners and their service station operators there exist two forms of agreement. Regarding the first, covering the supply of petroleum products, little need be said. Subject to cancellation on short notice, the supply agreement generally requires the major to provide, and the dealer to purchase, the latter's requirements of gasoline from the former. [2] The second form concerns the operation of the station property. [3] Although all majors distribute through stations operated by salaried employees, only four appear to use

---

1. In the Census of Business an establishment is classified as a service station only if over half of its gross receipts are derived from the sale of petroleum products. Thus an additional 100,000 establishments may dispense gasoline but are not service stations.

2. Many of these stations are supplied through independent jobbers or wholesalers. These stations are not considered in this study.

3. For a more detailed description see J. G. McLean and R. W. Haigh, The Growth of Integrated Oil Companies (Boston: Division of Research, Graduate School of Business Administration, Harvard University, 1954), pp. 476-78. A somewhat different classification is in S. N. Whitney, Antitrust Policies (New York: The Twentieth Century Fund, 1958), Vol. I, p. 126.

salaried operations for more than 50 or 100 stations.[1] The other majors use this method only in stations in which new dealers are trained, where parkway or turnpike commissions require, or in which new products and merchandising techniques are developed. About half of the 206,000 station properties are leased to the station operators by the major refiners. The remaining stations[2] are independent of the majors entirely, except for the product supply. Of these 100,000 or so leased stations the majors themselves own one third to one half; the remaining one half to two thirds are owned by third parties, leased to the majors on a long term basis for sublease to the operators.[3] For these leased stations the refiner-dealer agreements cover both the property lease and the product supply; the contracts are written coterminously. Generally the sales per station are greater at leased (and salaried operated) stations than at the independent stations.[4]

The lease agreements usually extend for one year, subject to renewal for an indefinite number of one year terms if both the refiner and the dealer are satisfied. No reason need be offered, however, if either should decide not to renew. Three and five year leases are being offered to "qualified" dealers, those whom the refiner considers desirable retail representatives for his brand. Additional common provisions of the leases include requirements that the dealer diligently promote the refiner's products, that the dealer not permit illegal or dangerous activities at the station, and that the dealer keep his station open for service either when competing stations are open or for specified hours during the day or evening. Rental provisions are flexible adaptations of two elements: a flat rent and a gallonage rent, for example, $250 per month plus 1 1/4¢ per gallon of gasoline sold. Variations are common in monthly differences in the flat rental, maximum and minimum dollar amounts, and exemptions in the gallonage rental for quantities below (or above) certain specified amounts.

---

1. Cities Service, Standards of California, Ohio, and Kentucky. The two most intensive users of this method, Standard of California with 20% and Standard of Ohio with 10% of their outlets, may prefer this method in areas of high population density and concentrated marketing.

2. Plus, of course, those outlets not so classified by the Bureau of the Census.

3. Some stations are leased to the major and subleased back to the owner-operator. Bank financing is thus facilitated for the operator. Sometimes the majors lease the land and then construct the station building.

4. In New Jersey in 1955, for example, the eleven majors in the aggregate used these methods in the following proportions: employee operated, 2.3%; leased, 37%; and independent, 60%. Of total gasoline sales in the state the proportions were respectively 3%, 55% and 41%. "Gasoline Price War in New Jersey," Report of the Select Committee on Small Business, U.S. Senate, 84th Cong., 2d Sess., Report 2810, July 26, 1956, p. 10.

Gasoline sold by these branded station operators is supplied by the major refiners. No major oil company, however, is vertically integrated into the manufacture of the non-petroleum products which are essential to the existence of most service stations; instead these tires, batteries, and accessories—TBA—are supplied by tire manufacturers to branded stations under two types of agreements with the sponsoring oil company. Under the purchase and resale plan a major oil company purchases from a tire firm the TBA for resale to the station operators. The TBA are manufactured to the specifications of the major; tire production generally utilizes the molds belonging to, and bearing the brand of, the oil company.[1] Sixteen of the largest 25 oil companies use this method; they undertake the warehousing and distribution functions of a wholesaler, and there is no direct contact between the tire manufacturer and the individual dealers. In 1958 12% of the replacement tires, or 8.4 million tires, were sold by this method.[2]

An additional 13.5% of the replacement tires, or 9.5 million tires, were marketed by the second method, a sales commission plan by which the tire manufacturers sell directly to the operators of the refiner's branded stations. Eleven of the 25 majors utilize this method.[3] The major sponsors the tire company's brands of TBA to the station operators; the tire company undertakes all the manufacturing and wholesaling activities, including the warehousing, distribution, and billing. Payment for the merchandise is made directly to the tire company. In return for this sponsorship—urging the dealers to stock and sell the chosen TBA brands—the major receives from the tire company a sales commission or "override" of ten per cent of the total revenue from TBA sales to the dealers.[4] This payment is substantially all profit for the oil companies, since the marginal cost of sponsorship is quite low. The profitability of this arrangement is indicated by the override receipts of three majors: in 1956 Texaco received $5 1/2 million from Goodrich and Firestone; in 1955 Atlantic received $1 million from Goodyear and Firestone, and Shell received $3 1/2 million from Goodyear and Firestone.[5]

---

1. For example, tires and batteries dispensed by Gulf, Mobil, and five Standard Oil Companies (Indiana, New Jersey, California, Kentucky, and Ohio) bear the Gulf, Mobil, and Atlas brands respectively.

2. The data of this and the following paragraph come mostly from two sources: National Petroleum News, Factbook Issue, Mid-May 1961, pp. 188, 189, and U.S. Congress, Senate, Select Committee on Small Business, Hearings, Dual Distribution in the Automotive Tire Industry—1959, 86th Cong., 1st Sess., 1959, pp. 366, 367.

3. Union of California and Sun use both methods.

4. 7 1/2% on sales to jobbers or wholesalers.

5. Federal Trade Commission Dockets 6485 (Initial Decision, p. 8), 6486 (Opinion p. 14), and 6487 (initial decision, p. 7) (all mimeographed). These dockets involve proceedings against the sales commission agreements between

The parties to these sales commission agreements do not limit themselves to one other firm. Among the tire firms Goodyear supplies its branded TBA to eight major oil companies by this method, Firestone seven, Goodrich six, and U.S. Rubber three. Four major oil firms have sales commission agreements with three tire manufacturers each, and four other oil companies, with two tire firms each.

That TBA sales are essential to the profitable operation of most service stations is indicated by the contribution to gross profit which these products make. Gross profit, the difference between sales revenue and the dealer's cost of goods sold, differs among the various products. For example, the dealer's gross profit from the sale of gasoline is 15-20% of sales, or about 5¢ per gallon, while the gross profit from most TBA is 35-40% of sales, and the gross profit for lubrication service approaches 100%.[1] The following figures indicate the proportions of total sales and of total gross profit for various products sold at average service stations:[2]

	Proportion of total sales	Proportion of gross profit
Gasoline	70-75%	50-55%
Lubricating oil	4-5%	7-8%
Tires and tubes	4-5%	4-5%
Batteries	1%	1%
Accessories	7-8%	10-11%
Miscellaneous, including lubrication service	7-8%	20-21%

The major oil companies attempt to induce their dealers to push the sales of TBA and oil. Each major has established ratios for the sales of each

---

Goodrich and Texaco, Goodyear and Atlantic, and Firestone and Shell, respectively. The Goodrich-Texaco docket was remanded to the hearing examiner, and the decisions in the other two proceedings, prohibiting the sales commission agreements, are being appealed by the respondents through the Federal Courts.

1. All the dealer's cost, other than the cost of goods sold, must be met from this gross profit: rent, electricity, heat, labor, insurance, etc. Gross profit is used here to avoid the arbitrary nature of assigning these costs to specific products. The return to the dealer, net of all costs except the opportunity costs of his own labor and invested capital, is 5-7% of gross revenue or sales. Most dealers do not get rich from their occupation.

2. These percentages are derived from data gathered from stations in Washington, D.C., and Baltimore, Maryland, by Fred S. Harris Co., Management Consultants, from data supplied by a major oil company, and from data gathered by the National Cash Register Co.

of these products in relation to the sales of 1,000 gallons of gasoline. These ratios are used as rules of thumb objectives which might be attained if the dealer does a "good selling job." Seldom are these ratios attained in practice, however, even at the majors' employee operated stations. The variability among station locations and operators apparently causes much variation among the actual ratios attained at stations of any particular sample. As a result, gasoline sales and these ratios are unsatisfactory predictors of actual TBA sales at any particular station. [1]

On the other hand, sales of gasoline at particular stations can be predicted with reasonable accuracy. The techniques vary from educated estimates by marketing experts to sophisticated estimates using formulas based on traffic density and speed, neighboring stations, shopping centers, and homes, and visibility and attractiveness of the station. Each major has some technique in use; the primary purpose of such estimates is the justification of capital expenditure on new station properties.

The Standard Stations Case (and subsequent legal activity) has caused the majors to eliminate all exclusive dealing clauses from their lease and sales contracts with their dealers: no interstate refiner-dealer contract today requires explicitly that a dealer handle particular product brands to the exclusion of competing brands. [2] Major oil companies insist that exclusive dealing as well as exclusive contract clauses have disappeared, and that the lessee dealers are independent businessmen, free to choose TBA and oil brands as each sees fit.

Except for the elimination of these exclusive clauses the refiner-dealer contracts have not been changed substantially from those existing at the time of the Standard Stations Case. [3] Incentives for exclusive dealing exist, and pressure on dealers to carry certain brands remains. But exclusive dealing is not an absolute concept: between the complete exclusivity—100% adoption of sponsored brands—and complete freedom of choice of brands by the dealers

---

1. For four samples of stations the coefficients of determination (r^2) between gasoline sales and TBA sales were 0.31, 0.06, 0.10, and 0.64. The r^2 between gasoline and lubricating oil was somewhat higher: 0.67, 0.61, 0.63, and 0.82. A substantial portion of the variance of TBA (and oil) sales is thus "unexplained" by gasoline sales. These four samples are, respectively, the samples of Tables 1, 3, and 4, and the employee operated stations of another major.

2. An intrastate exception in 1959 was the Cutler Oil Company of Lansing, Michigan, a distributor for Leonard Refineries. U.S. Congress, House, Select Committee on Small Business, Hearings before Subcommittee No. 5, Small Business Problems in the Petroleum Industry, 86th Cong., 1st Sess., 1959, pp. 210–11.

3. Standard Oil Company of California v. United States, 337 U.S. 293, the Record pp. 1518–1647 contains copies of the sales and lease forms used by Standard's six major competitors in the seven state Western area. Comparison of these forms with those in use today indicates no other substantial change.

lies a wide range in which some pressure can be brought to bear by the majors, some influence exerted, to cause the dealers to choose the sponsored brand rather than another. If the "pressure" is a lower price, then exclusive dealing loses its meaning. Although sanctions may only rarely be used, all that is required for exclusive dealing pressure to have effect is that a dealer think the refiner might not renew his contract or might fail to keep his gasoline tanks filled. Within this range, and short of the degree of exclusivity which the majors feel would incur the wrath of antitrust enforcers, most majors urge dealer acceptance of sponsored brands. There is wide variation among the majors in their actual practice of putting pressure on their dealers, just as within a particular firm the sales representatives differ in their methods of influence. But evidence of some pressure persists.

The ten per cent sales commission provides strong incentive to urge dealer acceptance of the sponsored TBA brands: the marginal productivity of sales effort by a major is quite high. The Federal Trade Commission in the sales commission proceedings against Firestone and Shell and against Goodyear and Atlantic found that such pressure existed despite lack of specific contractual provision for exclusivity. That the F. T. C. found no dealer whose lease was canceled solely for not purchasing sponsored TBA is not conclusive: the Department of Justice in the Standard Stations case found none either. Nor need cancellation for exercise of free choice be specifically found: the major can use other reasons for cancellation—not "promoting diligently" or not maintaining clean restrooms or grounds—and for non-renewal of the lease the major need not provide an excuse at all.[1] Majors' salesmen, in constant touch with the dealers, are paid not on a commission basis for products sold to dealers, but on similar schemes: bonuses and promotions for meeting or exceeding sales "objectives," often set in terms of TBA as well as petroleum products.[2] Moreover, the majors have discovered that "double-teaming," the simultaneous solicitation of a dealer's TBA purchases by sales representatives of both the tire company and the oil company, is a very effective method of gaining acceptance of the sponsored brands.[3]

On the other hand contractual and complete exclusive dealing, except perhaps in intrastate commerce and in employee operated stations, has entirely disappeared. Dealers canceled for failure to adopt sponsored brands

---

1. The turnover among lessee dealers is about 25% annually, that is, one of four lessee dealers terminates (or has terminated) his dealership each year for some reason. Dealer Turnover at Service Stations, Reports 1-6, American Petroleum Institute.

2. This emphasizes the "salesmanship" aspect of the pressure on dealers. However, often the sales representative has direct responsibility for keeping his stations, of which he has 15-25, supplied with dealers. Dealer terminations may not be in his interest.

3. Federal Trade Commission Dockets 6486 (Opinion, pp. 41-45) and 6487 (Opinion, pp. 24-28) (mimeographed).

are impossible to find. In fact good dealers, the majors claim, are hard to find, and often the majors have difficulty in keeping their stations in operation. However, the inference remains, that there exists pressure on dealers to stock and sell brands which otherwise they would not handle.

3. Interpretation

Complex relationships often require complex interpretations. The intricate form of vertical integration which the major oil refiners have developed "downstream" with their dealers can be understood by four complementary explanations. Together they contribute to an understanding of the motivations of the majors.

A. Cost saving

The lessee method of operating service stations minimizes the costs of distribution which the major must bear. Salaried managers may be subject to limitations on the hours per week which they may work, both in number and distribution throughout the week. Employees are more prone to the organizational efforts of labor unions. A lessee operator is not subject to these limitations, and the lessor thus can market his products at a lower distributional expense. A station which is open from 8 a.m. until 10 p.m., six days per week and sometimes on the seventh as well, is costly to operate by employees if they must cease work at the end of 40 or 48 hours. The average work week for lessee operators is about 70 hours. In addition the social security and unemployment taxes, to which an employer is subject, may be avoided by small dealers.[1] A study by one major oil company indicates that the operating costs of salaried stations are about one cent per gallon of gasoline greater than for leased stations. A 20,000 gallon per month station, not large by industry standards, thus costs $2,400 per year more to operate under salaried management than under lessee operation.[2] It is not surprising, then, to observe that only infrequently do the majors use salaried managers in their stations. When they do, the reasons generally involve training of personnel and development of new products and marketing methods. On the other hand, this explanation is not consistent with the previously noted practice of several majors, Standards of California and Ohio in particular, which use salaried managers in a substantial portion of their stations. Perhaps the concentration of marketing effort in centers of high population density explains this policy, which implies similarity in costs for these majors in salaried operations and lessee operations.

1. In the middle 1930's the movement from employee operations to lessee operations was given great impetus by the institution of state chain store taxes. Not only were these taxes heavy, in some instances they increased progressively as the number of stores increased. To avoid these taxes the refiners fired their salaried managers and leased the same stations back to them as lessee dealers.

2. Lessee dealers as independent businessmen may work harder for themselves than they would as employee-operators.

B. Guaranteed outlets for petroleum products

Executives in the oil industry insist that vertical integration into retailing is prompted by their interest in assuring outlets to final demand for refined products. Long term leases on many station properties prevent large scale movement away from a particular brand. Sponsorship of a TBA brand, in addition, comes from a sincere desire to serve retail customers with good products at the major's branded service stations.

Certainly much evidence is consistent with this interpretation of the refiner-dealer relations. The interests of any manufacturer include assurances that he can reach the market as well as consumers' satisfaction with all the products (not just the petroleum products) obtained under the refiner's brand. Quality of these non-petroleum products can be maintained by TBA sponsorship. Moreover, the stations over which the majors exercise some form of property control are the better stations in terms of sales of all products. In addition, there is some evidence that the majors at times subsidize the operation of three party leased stations: the rental receipts from the dealers fall short of the rental payments to property owners.[1] Such subsidization is reasonable in terms of acquiring assured outlets.

C. Returns for privileged access

The benefits to oil companies of sponsoring to their dealers certain non-petroleum products is apparent from the size of the sales commission payments, considered either as a percentage of sales or in aggregate dollar terms. Tire manufacturers are willing to pay handsomely for the delivery of a substantial block of stations, all of whose operators are under some pressure from their major oil company to purchase their TBA from the favored tire supplier. By such sponsorship the major is able to share in, perhaps to extract entirely, the benefits from the favored position which the (partial) exclusion of other TBA suppliers may provide. The purchase and resale plan may be similarly interpreted; the oil companies reserve their branded stations for themselves, and the profit on the sale of the sponsored TBA products is the return which the

---

1. In 1956 Esso received $12,225,026 while paying to the station owners $11,999,408 for the same stations. U.S. Congress, House, Select Committee on Small Business, Hearings before Subcommittee No. 5, Distribution Practices in the Petroleum Industry, 85th Cong., 1st Sess., 1957, p. 396. At least one major fares considerably worse than this in its attempt to effect a "washout" of these lease rental payments: in 1960 one medium sized major collected from its dealers only about half of its payments to station owners in the Eastern portion of the country. For its dealers in business less than one year, rental receipts were 54% of rental payments; for its dealers in business more than one year, the percentage was 51%. This tends to substantiate Spengler's prediction that these rental payments will accrue ultimately to the land (station) owner. Joseph J. Spengler, "Monopolistic Competition and the Use and Price of Urban Land Service," The Journal of Political Economy, LIV, pp. 385-412 (October 1946).

exclusion of other sellers affords. In other words, the sponsorship of TBA brands is profitable for the sponsor.

An explanation for the tire companies' interest in "buying" a group of stations may be suggested. Previous discussion [1] indicated the importance of joint production in service station operation: less than 60% of the gross profit from station operation is derived from the sale of gasoline and oil, while much of the remaining gross profit is contributed by the sale of TBA supplied by the tire companies under their agreements with the oil firms. The adoption of a brand of TBA (including a major's own brand) presents difficulties for the small tire companies which exist or plan to expand by selling to service stations but which lack the large scale distribution and warehousing facilities necessary to serve all the needs of a major oil company. Under the sales commission plans only the well accepted brands supplied by the tire firms with extensive distribution facilities have been chosen by the majors. Since the majors apparently prefer that one (or perhaps two) brands be sold throughout large segments of their marketing areas, rather than a variety of brands, the only tire companies able to meet their needs are the largest four or five. The sales commission payments, then can be interpreted as the payment by the tire companies for the erection of a barrier to the entry of new manufacturers and the expansion of existing firms in the industry. Without these barriers other tire producers or TBA suppliers would be able to nibble at the fringes of the service station market, injecting price competition for the patronage of the individual station dealers. The alternatives to the station operators are thus limited to the sponsored brands insofar as the sponsorship or exclusivity is effective. [2]

The sales commissions (or profits from TBA sales) must be sufficient to induce the oil companies not to manufacture their own TBA products. In addition the group of stations under a single refiner's brand must be considered a substantial group of outlets with regard to TBA sales. If this were not the case, the tire companies would be uninterested in making a sizable payment for the sponsorship in an unimportant segment of the TBA market. In addition the benefits to the favored tire companies of this barrier to other TBA suppliers may reflect the superiority of these sales techniques to other forms of advertising. A tire company under both the sales commission and purchase and resale plans has his products stored, displayed, and actively merchandised by interested dealers; no substitute TBA brand is available for the dealers to

1. p. 14 supra.

2. The Federal Trade Commission took explicit recognition of this barrier in its Goodyear-Atlantic and Firestone-Shell opinions (dockets 6486, pp. 57-59, and 6487, pp. 35-37, 42, mimeographed). Tire manufacture is highly concentrated: the largest four firms (Goodyear, Firestone, Goodrich, and U.S. Rubber) supply 79% of the total value of shipments in the country, and the largest eight, 91%. The sales commissions (1) permit a tire company to have access to a refiner's stations, and (2) keep the oil companies from manufacturing their own TBA.

distribute, hence the dealers' sales effort is directed toward the sponsored brand.

It should be noted that this explanation does not involve a barrier to the entry of petroleum refiners or station operators. With regard to the entry of firms into a particular industry, vertical integration in that industry affords no barrier (with one exception, noted below) which does not exist without that integration.[1] If entry is profitable on one level of the industry, then entry should be profitable and attractive on two or more levels at once. In the petroleum industry the integration of the refiners into the marketing level should afford no additional barrier to entry into either refining or marketing. However, entry into two levels demands more capital than entry into one level. If capital is rationed to the potential entrant, then the additional capital requirement may indeed bar his entry.

On the retail level there is ample evidence that the majors actively assist potential station operators into the industry. The costs of modern station construction, estimated as $50,000 at a minimum, effectively bar many potential dealers, who are generally regarded as poor risks by lending institutions. Some operators have difficulty in obtaining the necessary funds for working capital and for stocking a station with hand tools and an inventory of products.[2] The vertical integration of the majors, by use of the property leases, permits the dealers to obtain the necessary financing either through loans or consignment of inventory from the major or loans from banks or other sources with the major as guarantor.[3] This type of assistance is available, of course, only for a dealer who will sell the major's branded products, including the sponsored TBA brand. Entry into the retailing level by an unbranded or off-brand station operator apparently is not hindered by the activities of the majors. There is no indication that the majors singly or collectively have tied up all the desirable station sites; and the success of the private brand marketers suggests rather free entry.[4] In addition a guarantee of product supply at

1. Robert Bork, "Vertical Integration and the Sherman Act: The Legal History of an Economic Misconception," The University of Chicago Law Review (Autumn 1954), pp. 154-201, esp. pp. 194-96.

2. As a minimum several thousand dollars are necessary. The larger the station the larger is the requirement for these funds. Esso has estimated the requirement at $8,000 to $10,000. U.S. Congress, House, Select Committee on Small Business, Hearings before Subcommittee No. 5, Small Business Problems in the Petroleum Industry, 86th Cong., 1st Sess., 1959, p. 307.

3. U.S. Congress, House, Select Committee on Small Business, Hearings before Subcommittee No. 5, Distribution Practices in the Petroleum Industry, 85th Cong., 1st Sess., 1957, p. 240.

4. William A. Romeiser, "Competition Between Private-Brand and Major Brand Gasoline Retailers with Emphasis on Some Recent Federal Trade Commission Decisions," (unpublished Master's dissertation, School of Industrial Management, Massachusetts Institute of Technology, 1960), pp. 14-16, 21-24, 31-34.

wholesale can be gained from numerous majors, many of whom have had excess refining capacity recently.

On the refining level Bain has suggested that the barriers to the entry of new firms are "substantial," a classification between "very high" and "moderate to low."[1] In arriving at this evaluation no mention is made of the possibility that vertical integration into service station operation might be a requirement for entry into the refining level. Other authors have emphasized barriers to refining entry other than the availability of, or lack of access to, service station outlets.[2]

D. Rental maximization on leased property

The fourth reason for the refiner-dealer relationship involves the setting of the rental price on the station which maximizes the return to the refiner as station lessor. In other terms, if a particular station occupies a desirable location and as a result it sells a large quantity of, for example, gasoline, the lessor attempts to share in the benefits of this station superiority by setting the lease price accordingly: the rental payment which the dealer makes would be high. In view of the widespread use of gallonage rentals, it is not surprising that the rental payments reflect the majors' sharing in the benefits of large gasoline sales at particular stations.

On the other hand, if a station is superior not with respect to gasoline but with respect to a non-petroleum product, for example, tires, then the major as lessor would be induced to institute a similar profit sharing scheme, by which the major would appropriate a portion of the profit which otherwise would accrue to the dealer. But on what basis would the "proper" rental price be established? Surely the rent should depend directly on the volume of tire sales just as the rent does depend on the gasoline sales. However, tire sales may be information which the major does not possess, since the dealers (without exclusive dealing) may obtain their non-petroleum products from other sources. Under some conditions inspection of the dealers' accounting records is deficient for at least two reasons: dealers are notoriously poor accountants, and even if their records were perfect, many opportunities exist for keeping two sets of books; and inspection of the records of individual dealers would destroy the image of dealer independence which the majors so carefully cultivate. The sales commission and purchase and resale plans of TBA distribution permit the majors to obtain accurate information on sales by individual dealers, accurate to the extent that exclusivity is effective. Under the purchase and resale plan the information is already in the major's hands; under

1. Joe S. Bain, Barriers to New Competition (Cambridge: Harvard University Press, 1956), pp. 169, 170.

2. M. G. de Chazeau and A. E. Kahn, Integration and Competition in the Petroleum Industry (New Haven: Yale University Press, 1959), pp. 350, 352, 356, 357, 484–528; D. C. Hamilton, Competition in Oil: The Gulf Coast Refinery Market (Cambridge: Harvard University Press, 1958), p. 88.

the sales commission scheme, the TBA suppliers report to the majors the purchases of TBA by individual dealers.[1] Since this reasoning is applicable to the sale of any product through a station—a station may be superior with respect to any or several of its products, not just gasoline and tires—the rental payment may be adjusted to permit the major to share in the benefits of the station's superiority, for all products which show evidence of this superiority by high sales.

Several considerations are necessary for this explanation of the refiner-dealer relationship to be plausible. First, the non-petroleum products must be important contributors to the value of particular station sites. Almost half the gross profit from the operation of a service station comes from the sales of products other than gasoline. Second, if there exists a high correlation[2] between gasoline sales and, for example, tire sales, then the information derived from exclusive dealing and the TBA supply agreements with the tire firms would be superfluous: gasoline sales could be used as a proxy to estimate tire sales through individual stations, and a rental payment based on gasoline sales alone would accomplish the same result as a rental payment based on both tire sales and gasoline sales individually. Third, the same end could be attained if the wholesale prices on individual products could be adjusted to take account of these differences in station superiority. Proper prices to each dealer for each would accomplish much the same purpose. Perhaps fear of prosecution under the Robinson-Patman Act prevents this procedure. Probably more important, manipulation of individual product prices to individual dealers, although potentially workable, is not a practical alternative to the adjustment of lease prices.

The ultimate recipient of these benefits depends upon the source of the differences in sales of the various products. If the dealer himself is a superior dealer, perhaps because he can develop personal allegiances among his customers, then presumably he will receive the benefit from his ability. He would benefit regardless of the brand of petroleum products which he sold; his opportunity cost (to himself as well as his lessor) is greater than for a relatively "poor" dealer. This only modifies, but does not vitiate, this explanation, since other sources of superior sales exist. If the station location is the source of these superior sales, perhaps because of dense traffic and few competing stations, then the station owner would extract the benefit. A three party lease arrangement, under which the refiner is both lessee and lessor, merely means that the rental payment is syphoned from the dealer through the major to the third party as owner.[3] If the major is the station

---

1. The form of this information varies among refiners: sometimes the TBA are aggregated for each station, sometimes not. See Tables 1-4 infra.

2. The issue is not one of statistical significance. The point is that there is substantial portion of the variance in TBA sales which is "unexplained" by gasoline sales. This the data show. See footnote 1, p. 230.

3. See footnote 1, p. 233 supra.

owner, this question does not arise. Regardless of whether the major owns the station or leases the property from its owner for sublease to the dealer, the major as lessor is interested in maximizing the rental return which he receives. If the petroleum brand is the source of superior TBA sales, then the rental maximization explanations prevents the dealer from getting a free ride on consumers' preference for the lessor's brand, a preference which may spill over into the TBA brand choice of the consumer. These three sources of superiority, then, imply that the benefits of superiority must be divided between the station owner, the major, and the dealer.

One additional item must be noted. The total return to the major from TBA sales (the sales commissions or the profit under the purchase and resale plan, plus the station rentals) might be greater if the refiner would permit free choice of TBA brands to each of his dealers: presumably each dealer himself knows at least as well as the major which brand would "sell best" at the dealer's own station. With free choice, however, the major's actual return might be less than this potential return. This potential return is not practically attainable for the major: the information on sales of individual products through individual stations would not be available, and the major would be unable to get any return for TBA brand sponsorship.

4. Data on Rental Payments

Although explanation 3D regarding the setting of lease prices on service stations is reasonable, there may be some question whether refiners actually behave in this manner. Some refiners, when questioned, deny that they do. On the other hand, an executive of the Standard Oil Company of California testified before a Congressional Subcommittee in 1960:

> In the study that preceded our new lease we came to the conclusion that a rent based on gasoline volume alone has become outmoded. As the sales and services offered at service stations have become diversified, gasoline volume alone has ceased to be an accurate measure of rental value. . . . We now use a so-called overbasic rent which applies to sales and services beyond the usual range. [1]

The provision for overbasic rent, written into section five of Standard's lease,

---

1. G. C. Briggs, General Manager for Retail, Western Operations, Standard Oil Company of California, in U.S. Congress, House, Select Committee on Small Business, Hearings before Subcommittee No. 5, Small Business Problems in the Petroleum Industry, 86th Cong., 2d Sess., 1960, p. 674. R. H. Scholl, Vice President of the Esso Standard Oil Company, testified in a similar but less specific manner: "In trying to appraise the situation [at a particular station], we look at this outlet as a going business, and try to figure what kind of rent we reasonably ought to recover from that." U.S. Congress, House, Select Committee on Small Business, Hearings before Subcommittee No. 5, Distribution Practices in the Petroleum Industry, 85th Cong., 1st Sess., 1957, p. 396.

provides that the dealer pay a

> percentage of the income from sales and services in addition to gasoline, oils, tires, batteries, lubrication, tire and battery services, or sales and services of accessory items of the type which we made available to our dealer. [1]

Explanation 3D regarding rental payments can be expressed in this manner: major oil companies require exclusive dealing of their lessee dealers (to the extent that the majors feel possible under existing legal constraints) to carry sponsored brands of petroleum and TBA products, because information gained concerning the sales of various products at specific station locations may be used to adjust the lease prices, thus to reflect variations in the sales of individual products among the stations.

Both methods of TBA supply, the purchase and resale plan and the sales commission plan, result in the equivalent of varying individual rental payments among station operators. The sales commission scheme pays an oil company a constant percentage (10%) of the dealer's purchases of sponsored products, hence differences in quantities purchased by the various dealers cause different commission payments to the oil company. If the dealer occupies a "superior" station for TBA sales, then he purchases a large volume of these products from the favored tire company, and the override payment to the oil company is proportionately large. Small TBA purchases by a dealer likewise imply small dollar override. The purchase and resale plan can be similarly interpreted. If the percentage markup per unit of TBA is the same among all dealers, then the dollar profit derived by the major varies directly with the purchases of TBA by the dealers. There is nothing to guarantee that the sales commission rate or the chosen markup per unit is appropriate for all products or for all stations. The rental payments on leased stations can be adapted to "correct" an inappropriately established 10% or unit markup, if indeed the lessor thinks it should be changed.

For example, suppose

$$(1) \quad Y = a + \sum_{i=1}^{n} b_i X_i$$

where Y is the revenue to the oil company from selling particular products to a station operator, the X_i are the dollar sales (wholesale) of particular products to a station operator (and hence to final customers), and n is the number of products. Equation (1) is the sum of two equations (2) and (3):

---

1. U.S. Congress, House, Select Committee on Small Business, Hearings before Subcommittee No. 5, Small Business Problems in the Petroleum Industry, 86th Cong., 2d Sess., 1960, p. 684. The contract is reprinted in the same Hearings, p. 320-23 (1959).

$$(2) \quad Y' = a' + \sum_{i=1}^{n} b_i' X_i$$

representing the return from TBA sponsorship under both methods of TBA supply, and

$$(3) \quad Y'' = a'' + \sum_{i=1}^{n} b_i'' X_i$$

representing the return from adjusting the lease prices of individual stations.

With regard to equation (2), Y' is the sales commission payment (or the profit from sales of TBA) and the b_i are the sales commission rate, 10% (or the profit rates per sales dollar on TBA sales). Under both methods of TBA distribution the superiority of particular stations (for whatever product or combinations of products) is reflected in an increased return (Y') for the refiner.

Equation (3) reflects the extraction of additional return which the major appropriates by adjustment of the rental price of the station property, rather than by simple manipulation of the sales commission or profit rates. Y'' is the rental payment made by the dealer to the major as lessor, and the b_i'' represent the additional profitability made possible by the adjustment of the rental payment.

Thus

$$Y = Y' + Y'' = (a' + a'') + \sum_{i=1}^{n} (b_i' + b_i'')X_i$$

The oil companies are not indifferent to whether the return comes via b_i' or b_i''. Although b_i (as the sum of b_i' and b_i'') may remain constant for many combinations of these two components, the b_i'' can be adjusted only for those stations which the major leases to dealers. The b_i' greater than zero, then, do contribute to a major's income for those stations over which it has no lease control (and for which $b_i'' = 0$).

But do majors behave as equation (3) suggests? Regression equations have been fitted by the method of least squares to data of four samples of stations leased to dealers by three major oil companies. The notation for Y'' and X_i is the following:

RP	(Y'') Rental payment, dollars
G	Gasoline sales, thousands of gallons
MO	Motor oil sales, gallons
T	Tire sales, dollars
t	Tube sales, dollars
B	Battery sales, dollars
A	Accessory sales, dollars
TBA	Tire, tube, battery, and accessory sales, aggregated, dollars

Gr Grease sales, pounds (proxy for lubrication services)
I Investment in station property, by major oil company, dollars, original cost, undepreciated book value

The statistical significance of b_i'' can be interpreted thus: the profit per gallon of gasoline and the 10% override (or the major's unit profit on TBA sales supplied under the purchase and resale plan) provide an imperfect method of maximizing the return from particular stations, which differ in sales of particular products. If the estimates of b_i'' are not statistically different from zero, then the oil companies do not appear to use individual rental payments (Y'') to supplement the 10% override or the unit profit margin (Y').[1]

The variables used generally refer to annual rates for the year in which the data were available (1959 or 1960). Monthly data are indicated by a subscript (for example, RP_m) and refer to the mean monthly quantity during the year. Not all the variables (X_i) were available for each sample of stations, hence the specific regression equation varies among the samples. Since the data have been obtained from the oil companies rather than from the station operators, the data reflect only the sales of products supplied by the major (or by the favored tire firm) to the dealers. Purchases of X_i by the dealers from other sources are not included. This is not a damaging qualification, however, since the behavior of the majors in setting the rental prices of particular stations depends upon the information which they have, not necessarily on speculation about the sales volume of non-sponsored products.

Tabes 1, 2, 3, and 4 present the results of the regression analysis. The first equation in each of the tables regresses the rental payment (Y'') on all the variables (X_i) available for that particular sample. Additional regression equations for each sample have been computed to indicate what effects, if any, would result when certain variables were omitted from the estimation process. The eliminated variables generally were not significant at the 0.05 level.[2] The statistical significance of these coefficients[3] suggests that the major oil companies do behave as the rental maximization explanation

---

1. For majors using the sales commission plan $b_i' = +0.10$; since $0 < b_i = (b_i' + b_i'') < 1$, then $-0.10 < b_i'' < +0.90$.

2. Investment (I) was eliminated from equations 2 and 3 of Table 1 because of possible ambiguity of interpretation: does rental payment depend on investment, or does the investment depend on the (anticipated) rental payment? In Table 2 batteries (B) and tires (T) appear respectively in equations 3 and 4; individually they are significant, together, in equation 2, neither is significant.

3. The null hypothesis, that a particular coefficient of these tables is in fact zero, is rejected on a 0.05 level of significance if the ratio between the b_i'' and its standard error exceeds 1.96 (2.58 for the 0.01 level).

TABLE 1

Description of Sample

50 stations, selected randomly from the stations owned by a major oil company and leased to lessee operators in the oil company's New England Division. Data are for the year 1959. This major and the major of Table 2 are the same.

Source of TBA

Purchase and resale plan; oil company's private brand of tires and batteries.

Coefficients of multiple regression equations

Standard errors are in parentheses. Rental payment is regressed on the indicated variables in each equation. Rental payment is the mean monthly payment during the year.

Equation Number	Constant Term	G (Gas.)	MO (Oil)	T (Tires)	t (tubes)	B (Batteries)	A (Access.)	I (Invest.)	R^2	S_u
1.	+15.56	+0.612	-0.0153	-0.0110	+0.0878	+0.0572	+0.0882	+0.00142	0.849	87.59
	(34.16)	(0.155)	(0.0181)	(0.00644)	(0.136)	(0.0242)	(0.0255)	(0.000534)		
2.	+36.439	+0.622	—	-0.00582	—	+0.06918	+0.0878	—	0.820	92.36
	(32.035)	(0.119)		(0.00579)		(0.0251)	(0.0261)			
3.	+32.67	+0.624	—	—	—	+0.0606	+0.08105	—	0.816	92.37
	(31.82)	(0.119)				(0.0236)	(0.0252)			
Means: RP_m =	$368	340	2818	$2733	$121	$680	$1,011	$55,099		

Source: Computed from confidential data supplied by a major oil company.

TABLE 2

Description of Sample

50 stations, selected randomly from the stations leased by their owners to the major oil company and subleased by the major to lessee operators in the oil company's New England Division. Data are for the year 1959. This major and the major of Table 1 are the same.

Source of TBA

Purchase and resale plan; oil company's private brand of tires and batteries.

Coefficients of multiple regression equations

Standard errors are in parentheses. Rental payment is regressed on the indicated variables in each equation. Rental payment is the mean monthly payment during the year.

Equation Number	Constant Term	G (Gas.)	MO (Oil)	T (Tires)	t (tubes)	B (Batteries)	A (Access.)	R^2	S_u
1.	+20.44	+0.865	+0.0112	−0.00513	+0.00738	−0.0503	+0.0678	0.743	49.52
	(30.26)	(0.134)	(0.0108)	(0.00379)	(0.113)	(0.0286)	(0.0226)		
2.	+30.41	+0.8966	—	−0.00477	—	−0.0431	+0.0707	0.736	49.07
	(27.92)	(0.123)		(0.00366)		(0.0275)	(0.0204)		
3.	+20.62	+0.9557	—	—	—	−0.0554	+0.0581	0.726	49.45
	(27.09)	(0.115)				(0.0260)	(0.0181)		
4.	+36.43	+0.8573	—	−0.00674	—	—	+0.05702	0.721	49.84
	(28.08)	(0.122)		(0.00349)			(0.01869)		
Means: RP_m =	$281	253.5	2042.6	$1,701	$69	$385	$680		

Source: Computed from confidential data supplied by a major oil company.

TABLE 3

Description of sample

50 stations, selected randomly from the stations leased to lessee dealers by a major oil company throughout that major's entire marketing area (most of the United States). Data are for the year 1959. No distinction is made for actual ownership of the station properties.

Source of TBA

Sales commission plan. The method of reporting the TBA sales to the major the TBA sales made to the major's dealers does not permit a breakdown of TBA sales into its components.

Coefficients of multiple regression equations

Standard errors are in parentheses. Rental payment is regressed on the indicated variables in each equation. All variables are mean monthly figures during the entire year.

Equation Number	Constant Term	G_m (Gas.)	MO_m (Oil)	TBA_m (TBA aggregated)	R^2	S_u
1.	+21.321	+13.51	+0.0226	+0.03403	0.951	27.16
	(14.274)	(0.744)	(0.0699)	(0.01715)		
2.	+22.183	+13.69	—	+0.03491	0.951	26.89
	(13.884)	(0.478)		(0.01677)		
Means: RP_m =	$424.3	28.5	281.8	$333.4		

Source: Computed from confidential data supplied by a major oil company.

TABLE 4

Description of sample

270 stations selected randomly from the stations leased to lessee dealers by a major oil company throughout that major's New England Division. Data are for 1960. No distinction is made for actual ownership of the station properties.

Source of TBA

Sales commission plan.

Coefficients of multiple regression equations

Standard errors are in parentheses. Rental payment is regressed on the indicated variables in each equation. All variables are annual amounts. Tire variable includes tube sales.

Equation Number	Constant Term	G (Gas.)	MO (Oil)	T (Tires)	B (Battery)	A (Access.)	Gr (Grease)	R^2	S_u
1.	+425.6	+7.172	+0.254	+0.0623	-0.0656	+0.183	-0.371		
	(83.1)	(0.571)	(0.0659)	(0.0191)	(0.0881)	(0.0543)	(0.0763)	0.770	529.95
2.	+417.49	+7.122	+0.2257	+0.0488	—	+0.1959	-0.1949		
	(86.24)	(0.605)	(0.0821)	(0.0193)		(0.0533)	(0.2348)	0.746	553.88
3.	+412.18	+7.206	+0.1860	+0.0489		+0.1978			
	(85.96)	(0.597)	(0.0667)	(0.0193)	—	(0.0533)	—	0.746	553.56
Means: RP_m =	$2,582	211.5	2,050	$2,636	$391	$734	369		

Source: Computed from confidential data supplied by a major oil company.

argues, that rental payments are adjusted at individual stations in accordance with sales of individual products.

As expected, the coefficients for gasoline sales are significant and imply an additional $7 to $14 of rent [1] for a marginal 1,000 gallons of gasoline sales. Assuming a 5¢ per gallon margin for the dealer, common in the industry, about a fourth to a fifth of the dealer's gross profit from this sale is thus appropriated by the major through a higher rental payment. This result was anticipated from the use of a gallonage rental clause in the lease agreement.

Motor oil sales, significant only in Table 4, imply a marginal rental payment of 19¢ for a marginal gallon of sales, or about a quarter of the approximately 80¢ gross profit for the dealer. The lack of significance of this variable in the other three samples might be explained by the high correlation between gasoline and oil sales: use of oil sales might provide no additional information for the refiner in the estimation of rental payment, thus oil sales (in this sense) would be superfluous.

The coefficient of TBA (aggregated) in Table 3 implies that the 10% override (Y') is "too low" by 35%; the rental payment (Y'') is increased 3.5¢ for a marginal dollar of TBA sales to the dealer, thus extracting about one tenth of the resulting gross profit for the dealer.

The negative coefficient for tire sales in Tables 1 and 2 might be explained by the purchase and resale plan for TBA supply, which this major uses. The coefficients suggest that monthly rental is lowered about $0.006 for an additional dollar of annual tire sales, or −7.2¢ rent during the year for an increase in (wholesale) tire sales of one dollar. One might expect that the operators of large stations (those with large tire volume) might themselves individually be able to exert more bargaining power on their major for lower tire prices than the operators of small stations might be able to exert. However, fear of Robinson-Patman prosecution possibly precludes the major from varying the price of tires among stations, hence the major might give a lower lease price rather than a lower tire price to these large stations.

The coefficient of tire sales in Table 4 suggests that the 10% override is "too low" and should be 14.9%, that is, 10% plus 4.89%, for tire sales. This major raises the rental payment of the dealer 4.9¢ for the marginal dollar of (wholesale) tire sales, thus extracting about an eighth of the dealer's 40¢ gross profit.

Battery sales, significant in Tables 1 and 2, nevertheless have different signs. If the negative sign in Table 2 is a "mistake," [2] which the small annual mean sales ($385) suggests, then the positive sign of Table 1 indicates a marginal rental payment of 6¢ per marginal dollar of battery sales, or one sixth to one seventh of the dealer's gross profit.

---

1. For each of these samples the figures are $7.49 (12 x 0.624), $10.28 (12 x 0.857), $13.69, and $7.21 respectively. Rental payments for samples in Tables 1 and 2 are monthly, in Tables 3 and 4, are on an annual basis.

2. That is, if the coefficient is in fact zero.

In Tables 1 and 2 the coefficients of accessory sales imply marginal rental increases of 97.2¢ (12 x 0.081) and 68.4¢ (12 x 0.057) per marginal dollar of accessory purchases by the dealer. However, sales of these products provides the dealer with only 50-55¢ of gross profit. How can more than this amount be extracted by the major in the adjustment of the lease rental? Two interpretations are possible. (1) The wholesale prices of the products are low, and the return to the major is made up in higher rent; that is, the unit margins to the dealer are thus higher than the assumed gross profit figures. (2) The dealers who purchase large amounts of accessories may also be dealers who do a great deal of repair work. Even such minor repairs as installation of fan belts and replacement of spark plugs may result in labor charges and thus in higher gross profit for the dealer. The major thus may use the accessory sales as a metering device to estimate the dealer's repair work, thus to extract some profit from repair work in addition to the dealer's sales of accessory items.

The coefficient of accessory sales in Table 4 is more easily interpreted: a marginal dollar of sales results in 20¢ additional rental, about two fifths of the dealer's gross profit from sales of these accessories to his customers.

5. Summary

The explanations offered in this essay for the refiner-lessee dealer relationship imply conclusions which contrast with the blanket lessening-of-competition conclusion of the courts. Lower cost distribution and guaranteed access to final demand, as explanations for this relationship, imply no lessening of competition; they are quite consistent with competitive market behavior. On the other hand, erecting barriers to the access of alternative TBA suppliers, as the explanation regarding benefits for privileged access to branded lessee stations implies, can be interpreted as monopoly creating: the privileged tire firms may be freed from competitive harassment of other suppliers in a substantial portion of the market. However, additional information and analysis are necessary to determine if manufacturers are indeed "foreclosed" as the privileged access explanation implies. The rental adjustment explanation is consistent with the refiner's appropriation of producers' surplus,[1] which otherwise would fall to the dealer, and no supply restriction is apparent.

1. A. Marshall, Principles of Economics (8th ed., London: Macmillan and Co., 1952) Appendix H, p. 668.